When Group Work Doesn't Work: Insights from Students

Yunjeong Chang[†‡] and Peggy Brickman[§]*

[†]Department of Instructional Technology and [§]Department of Plant Biology, University of Georgia, Athens, GA 30602

ABSTRACT

Introducing group work in college science classrooms can lead to noticeable gains in student achievement, reasoning ability, and motivation. To realize these gains, students must all contribute. Strategies like assigning roles, group contracts, anonymous peer evaluations, and peer ratings all encourage student participation. In a class using these strategies, we conducted in-depth interviews to uncover student perceptions of group work in general and the utility of these support strategies. Students in both high- and low-performance groups still complained of unequal contributions while praising the social support provided by groups. Students who scored highly on tests were more likely to recognize the benefits of group work, regardless of their groups' overall performance levels, while lower-scoring students perceived group work as time-consuming "busy work" with little cognitive benefit. Comments from anonymous peer evaluations differed only subtly between high- and low-performance groups. Numerical ratings on these evaluations did correlate with overall group performance. However, students in lower-performance groups assigned harsh ratings to their low-scoring members, while students in higher-performance groups were more generous in their ratings for low-scoring members. We discuss implications of relying on support strategies for promoting productive group work.

INTRODUCTION

Science education policy (Handelsman *et al.*, 2004; American Association for the Advancement of Science, 2010; Couch *et al.*, 2015; National Research Council, 2015) advocates for including peer interactions (referred to as "group work" hereafter) in college courses because they provide opportunities for students to practice scientific reasoning, critical-thinking, communication, and problem-solving skills that have been shown to result in greater gains in achievement (Slavin, 1991; Springer *et al.*, 1999; Johnson *et al.*, 2000; Armstrong *et al.*, 2007; Preszler, 2009; Freeman *et al.*, 2014; Batz *et al.*, 2015). Group-work pedagogies with demonstrated evidence of effectiveness include collaborative learning (Phelps and Damon, 1989), cooperative learning (Slavin, 1991), team-based learning (Michaelsen *et al.*, 2014), peer instruction (Crouch and Mazur, 2001), SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs project) in physics (Beichner *et al.*, 2007), and POGIL (process-oriented guided-inquiry learning) in chemistry (Moog and Spencer, 2008). All of these group-work pedagogies encourage students to construct their own understanding of scientific concepts through a process of negotiation and consensus building with their peers (Solomon, 1987; Latour and Woolgar, 2013). Group work also provides a basis for social comparison, social learning, and social cognition (Solomon *et al.*, 2010), and students working in small groups may make gains in terms of achievement, motivation, and self-efficacy as a result of this comparison (Bandura, 2000; Hernandez *et al.*, 2013).

Several integrated theoretical frameworks have been espoused within educational psychology to explain the different constructs (motivational, social, and cognitive)

that influence the achievement effects of group work (Sweet and Michaelsen, 2007; Slavin, 2014). Social interdependence theory (Johnson and Johnson, 2009) is a particularly helpful theoretical framework, because it describes five major variables that mediate the effects of cooperation, including motivational, social, and cognitive aspects. The first variable described in social interdependence theory is positive interdependence: individuals' perception that they can reach their goals if and only if the other individuals with whom they cooperate also reach their goals and, therefore, promote one another's efforts to achieve the goals. The second variable is individual accountability: the responsibility to complete one's own share of the work and also facilitate others' work. The third variable is promotive interactions: individuals motivating and facilitating the work of others through sharing resources, providing help to one another, challenging reasoning and conclusions provided by group members, and taking varying points of view into account. The fourth variable is the appropriate use of social skills: skills in which individuals get to know and trust one another, communicate, support one another, and resolve conflicts that arise. Finally, group work should provide a mechanism for group processing and reflection: encouraging students to set collective goals, assess positive and negative group interactions, and provide feedback to group members. Instructors play a major role in promoting the variables required for these aspects of social interdependence during group work, and students express greater satisfaction with group work when their instructors include support strategies to assess and foster group collaboration (Chapman and Van Auken, 2001).

Recommended support strategies to foster effective collaboration include assigning roles, group contracts, peer evaluations, and peer ratings that measure differences in contributions.

Role Assignment

Assigning tasks or roles for students to assume while completing tasks is recommended as a way to promote individual accountability and ensure that instructors can monitor contributions (Chapman and Van Auken, 2001; Davies, 2009). Group work pedagogies like POGIL (Moog and Spencer, 2008) and SCALE-UP (Beichner et al., 2007) recommend assigning specific roles to promote critical discussion and to prevent students from either dominating discussions or avoiding conflict by accepting the quickest answer during problem-solving tasks (Heller and Hollabaugh, 1992). Role assignment has been shown to promote greater learning gains (Bailey et al., 2012) and student satisfaction (Brown, 2010). The caveat remains that supervision is required so that students cycle through the cognitive acts of listening and recalling that are necessary for greater exchange of ideas and thus learning (O'Donnell, 2006).

Group Contracts

Groups that discuss expectations before group work begins and draft group contracts that spell out consequences for failure to meet expectations foster appropriate use of social skills (Chapman and Van Auken, 2001; Oakley et al., 2004). Feelings of interdependence, cohesion, psychological safety, and confidence all strengthen the belief that investment in group activities will pay off and thus encourage students to engage in cognitive processes key to learning (Van den Bossche et al., 2006).

Peer Evaluations

Anonymous peer evaluations in which students reflect on their own and others' contributions and group dynamics and have the ability to inform the instructor of problems within the group help promote reflection, group processing, and individual accountability (Harkins and Jackson, 1985; Strong and Anderson, 1990; Brooks and Ammons, 2003; Oakley et al., 2004; Aggarwal and O'Brien, 2008). Evaluations also help reduce social loafing, a situation in which students in a group commit less effort to a group project, because they believe their lack of effort will not be identified, and free riding, a situation in which students knowingly allow others to complete their work for them (Aggarwal and O'Brien, 2008). In Strong and Anderson's (1990) study of student opinions about group work, students indicated that they believed that peer evaluations do reduce free riding, but they rated other factors—including group cohesiveness, small team size, the option to "divorce" a team member, or the option to leave a team—as having a stronger effect on reducing free riding. Students also rated the divorcing option as more effective at motivating team members than end-of-semester evaluations, which have been shown to be negatively associated with good team experiences (Bacon et al., 1999). This suggests that peer evaluations may encourage students to tolerate bad behavior, knowing they can exact retribution at the end of the semester. In addition, researchers have found that students may be unlikely to provide honest evaluations of their peers and are unlikely to directly confront free riders (Strong and Anderson, 1990).

Peer Ratings

There are challenges to determining the effort and achievement level of individual students during group tasks. On typical group tasks, all group members work together without being evaluated individually on the final product by the instructor, and all group members receive the same grade for the final product. If the level of contribution differs among group members, then the single group grade may not accurately reflect individual effort or performance. One solution is to collect information on the contributions of each group member using quantitative ratings and to adjust the final scores to reflect lack of effort (Latane et al., 1979; Bartlett, 1995). Students should be aware of the contributions of each member, and if they can be trusted to provide accurate ratings, the ratings could be used to derive a numerical weighting factor to adjust group grades accordingly. Researchers have recommended administering both holistic rubrics in which students are given points that they must divide between the group members to gauge contributions (Lejk and Wyvill, 2001; Johnston and Miles, 2004) and analytical rubrics with multiple indicators such as attendance, cooperativeness, and academic contributions to gauge contributions (Kaufman et al., 2000; Stefanou et al., 2001; Kilic and Cakan, 2006). Holistic rubrics have been found to be more effective at identifying very good and very weak contributors to the group, whereas analytic assessments are able to identify small differences in group contributions and may be more effective for providing formative feedback (Falchikov and Goldfinch, 2000; Lejk and Wyvill, 2001). Holistic rubrics, however, need to be adjusted by calculating an individual's contribution to maximize the correlation between actual group scores and true levels of a student's contribution (Zhang and Ohland, 2009).

TABLE 1. Overview of research questions and data sources

	Research questions	Data sources	
		Qualitative	Quantitative
1	Do students' overall perceptions of group work differ between high- (top 50% on rank-ordered group assignments and tests) and low-performance groups (bottom 50%)? And do perceptions also differ if they are high (top 33% using ranked final course grades) and lower scorers (bottom 33%)?	Interviews conducted with high- and low-scoring students in high- and low-performance groups	Group scores and individual test scores
2	To what extent do students' use group-based support strategies (e.g., roles, group contract, and peer evaluations) to communicate feedback to their peers?	Interviews conducted with high- and low-scoring students in high- and low-performance groups	Group scores and individual test scores
3	Do the frequency of students' reports of the five elements of social interdependence theory differ between high- and low-performance groups?	Analysis of categories and themes prevalent in peer-evaluation comments	Comparison of frequency of categories of comments on peer evaluations when administered midsemester (formative) versus at the end of the semester (summative)
4	To what extent do peer-evaluation ratings and comments distinguish high- and low-performance groups and higher- and lower-scoring students within those groups?	Comparison of comments left by peers on mid- and end-of-semester evaluations	Comparison of ratings left by peers on midsemester and end-of-semester evaluations

Faculty members have reported concern about providing sufficient support to ensure that students are interacting productively during group work in large-enrollment classes (Pundak and Rozner, 2008; Freeman and Greenacre, 2011; Barkley *et al.*, 2014). And many faculty attest to difficulties they have encountered implementing group work in their classes and have reported frustration with dysfunctional groups that require a great deal of supervision to ensure equitable student participation and reward productive social interactions (Kreijns *et al.*, 2003; Davies, 2009; Svinicki and Schallert, 2016).

In an attempt to improve student learning through group work, we implemented many of these recommended strategies for supporting group dynamics in a very large enrollment introductory biology course that was often linked to poor student learning, attitudes, and retention (Seymour and Hewitt, 1997; Barr *et al.*, 2008; Chang *et al.*, 2011; National Academies of Sciences, Engineering, and Medicine, 2016). In a preliminary study, we were able to demonstrate that group activities benefited students at diverse performance levels, but the students also reported persistent social loafing (Chang and Hannafin, 2015).

Managing group dynamics for successful implementation of group work is time intensive for instructors. We hypothesized that exploring students' perceptions of group work might provide insights into which time-consuming support strategies (e.g., role assignment, group contracts, peer evaluations at the midpoint and endpoint of the semester, and summative peer ratings) students were using effectively. We also hoped that the perceptions of students in high-performance groups could provide insight into critical features to be nurtured and enhanced in all groups. Because this was a primarily observational study aimed at characterizing student perceptions of group work and use of support strategies designed to monitor and improve group interactions in a large-enrollment college classroom, we employed a concurrent mixed-methods design (Creswell, 2009). We characterized student perceptions of group work using qualitative interviews and written comments submitted

on peer-evaluation surveys. We also compared the interviews, comments, and numeric peer-evaluation ratings given by and from students at different achievement levels, which we determined using quantitative data from course assessments (Teddlie and Tashakkori, 2003). Table 1 outlines our major research questions and the data sources used to address each question.

METHODS
Instructional Setting
We examined learning performance and group work for 246 students enrolled in an introductory biology course for non–science majors at a large public university in the southeastern United States. The course included two 75-minute classes per week and comprised five different content units. Class time was devoted to providing core content through instructor minilectures and daily individual and group activities that required students to apply the content to specific situations (Brickman *et al.*, 2012). Individual assignments included clicker questions, preclass written assignments to prepare for group work, practice tests, and a final unit test composed of multiple-choice questions. Group work included completing in-class worksheets to structure or organize content knowledge (e.g., drawing a diagram, finding relevant resources from websites) as well as outside-class group projects that required students to apply their knowledge (Brickman *et al.*, 2012). In addition, after completing unit tests individually, group members collaboratively answered the same test items again to earn a group test score. Group activities were designed to build social interaction and interdependence through deadlines that required individual work be submitted before beginning group work, initial activities that built rapport, and grading that emphasized attaining common goals (Deutsch, 1949). All group members received the same score for work from their group. All activities and test scores were weighted to compute each student's final course grade: individual test scores: 36%; clicker questioning: 12%; individual assignments to prepare for group work: 12%; group assignments and projects: 28%; and group tests: 12%.

Group Formation and Support Strategies

During the second week of the semester, the instructor asked students to organize themselves into groups of four or five members, resulting in 65 groups that completed group assignments and tests together throughout the rest of the semester. Owing to a 5% withdrawal rate, no five-member groups were present at the midpoint. The instructor did not dictate group composition because of a lack of clear consensus from the literature on the most effective characteristics to use when forming groups and because autonomous group composition minimized logistical and practical problems in course administration (i.e., tackling individual group requests) and students' resistance (i.e., request for changing groups or sitting near the front of the auditorium). Individuals assigned themselves to different roles within their groups for daily activities: manager/spokesperson, researcher, recorder, and whiteboard writer. The manager/spokesperson supported group processes during discussions (i.e., time management, soliciting ideas from all, speaking for the group); researchers gathered additional ideas for the group from class notes and the Internet; recorders summarized discussions and submitted group worksheets; and whiteboard writers created diagrams and figures on worksheets. At the beginning and during the semester, the instructor gave minilessons to remind groups to rotate roles in an attempt to balance participation by each group member (Johnson *et al.*, 1998). The instructor also created a section in the group worksheets where the assigned roles of each student in the group could be recorded.

Throughout the semester, groups were given opportunities to promote productive interactions and mediate conflict. At the beginning of the semester, groups created a contract in which they established expectations, set ground rules (e.g., penalties for failure to participate), and established communication channels. At the midpoint of the semester, groups were encouraged to revise and resubmit their contracts if necessary and to complete an anonymous online peer-evaluation survey to provide midcourse feedback to one another. Students also completed an online, end-of-course peer-evaluation survey to assess each group member's contribution. These surveys included eight numeric questions that asked students to evaluate individual group members' preparation, participation, collaboration, attitude, and performance during group work. For example, "Did this person participate in discussions during class? This could include sitting with the group during class, attending regularly, etc." Students ranked each team member on a four-point scale for each question (1 = unacceptable performance, I would fire this person; 2 = improvement needed; 3 = good, met or exceeded all expectations; 4 = outstanding, a rare individual). Text for each question in the survey can be viewed in Appendix A in the Supplemental Material. Each group member was finally asked to respond to one open-ended question: "Please provide written comments about each of your team members so they can learn how you viewed their contributions to the team. After all evaluations are submitted they will be able to read these comments but not tell who they came from. You and they can use this feedback to improve your future performance." Students were also asked to rate the quality of each group mate's contributions to the group on a scale of 0–100, similar to earning a grade in a class. Students were told that their group scores would be adjusted based on the average

rating that they received. The Opensource online platform that we used (Simple Team Experience Assessment Measure—STEAM) used the expected-contribution method (Lejk *et al.*, 1996), which is based on adjusting the group grade by adding a deviance from the expected contribution. If you have a group of four members, each member is expected to contribute 25% of the work. If a group member's average score indicates that he or she completed only 20% of the work, then the assigned group score is adjusted by subtracting 5% of the group score. Group members were able to view the scores and comments from their peers after all group members had submitted their responses.

Quantitative Data Collection and Analysis

Three different measures were used to compare individual students and groups. First, we calculated group performance levels based on the rank-ordered group assignment scores and averaged group scores on a series of group tests. We were interested in how groups performed relative to one another, so we rank ordered assignments within each unique unit, because topics differed and the variance of scores differed between units (Kruskal and Wallis, 1952). To have an adequate number of groups for statistical analysis, we ranked groups based on the median group score. Groups ranked above the median score were classified as high-performance groups ($n = 32$), while groups that scored lower than the median were categorized as lower-performance groups ($n = 33$). Categorizing groups into high-performance groups and low-performance groups allowed us to select balanced numbers of interviewees and compare the perceptions of students within those groups (more details about the distribution of the interviewers are described in the *Qualitative Data Collection and Analysis* section).

Second, we used final course grades to divide students into higher-, mid-, and lower-scorer categories. Students ranked in the highest 33% using ranked final course grades were categorized as higher scorers ($n = 82$), those in the middle 33% as midscorers ($n = 79$), and those in the lowest 33% as lower scorers ($n = 84$). This allowed us to quickly identify students to interview to address question 1, which compared perceptions of higher- and lower-scoring students within groups, and question 2, which addressed the use of group-based support strategies before the end of the course. A breakdown of the distribution of students at high-, mid-, and low-scoring levels is outlined in Figure 1. Eighty-seven percent of the groups contained a mixture of students of different scoring levels (heterogeneous), rather than containing students of similar scoring levels (homogeneous). All of the high-performance groups included at least one higher-scoring student, while only 10 of the 33 low-performance groups included at least one or two more higher-scoring students in their groups.

To test our research question 4, whether numeric peer-evaluation ratings served as a quantitative measure of students' perceptions of their group members, we used the average ratings that students received to compare average ratings between high- and low-performance groups as well as high- and low-scoring students within those groups. To investigate the effects of group performance, we used a random effects model that can control for the variance associated with random factors that may occur during the interaction between students and groups (Judd *et al.*, 2012). By using random effects for students
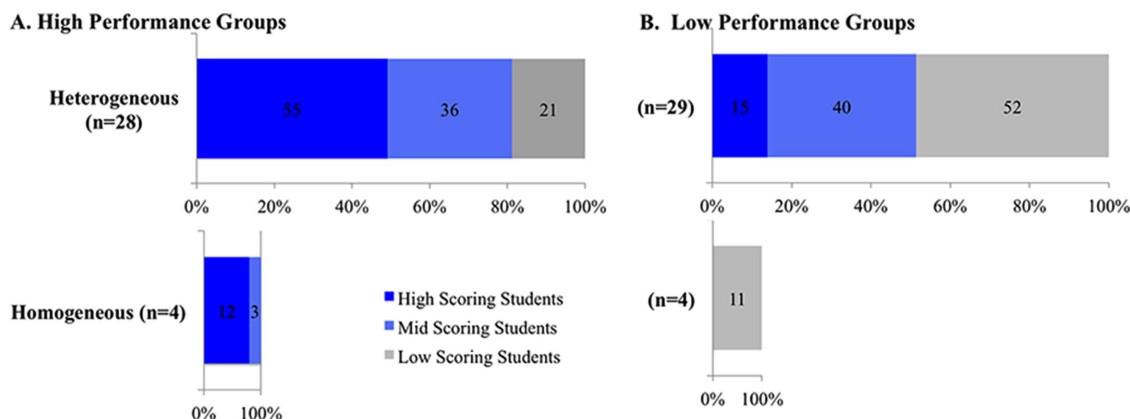
FIGURE 1. Most student groups, either high-performance (A) or low-performance (B), were composed of a mixture of high-, mid-, and low-scoring members and classified as heterogeneous (top row), 57/65 total groups. A much smaller number of groups (8/57) were homogeneous (bottom row) and were composed of either all high- (three groups), all mid- (one group), or all low-scoring students (four groups).

in different groups, we controlled for the influence of different student interactions associated with group variables. In our model, we used the score gap between group and individual scores as a dependent variable and group performance levels as the fixed effects. To determine whether a relationship existed between peer-evaluation scores and individual test scores, we used a paired t test. A student's peer-evaluation rating score was used as an independent variable, and individual test scores as a dependent variable.

Qualitative Data Collection and Analysis
We conducted in-depth interviews (Esterberg, 2002) to explore the range of students' perceptions, attitudes, and participation in group work in order to gain better insight into the briefer comments made on the peer evaluations. Appendix B in the Supplemental Material contains our interview protocol and question items. We recruited interviewees at the beginning of the last unit (unit 5). The first author (Y.C.) verbally recruited interviewees by announcing the interview in front of the class and sending out an email. The verbal and email announcements included information of the purposes and the foci of the study and the interview. From among 27 volunteers, the authors purposefully selected 15 interviewees to ensure that we had seven higher-scoring students and eight lower-scoring students based on their test scores. Among the seven higher scorers whom we interviewed, five were from high-performance groups and two were from low-performance groups. Among the eight lower scorers whom we interviewed, three were from high-performance groups and five were from low-performance groups. The number of high and low scorers selected for interviews was related to the proportion of these students in the respective groups (see Figure 2 for the group score distribution of the interviewers and Figure 1 for breakdown of student distribution in groups). Thirteen groups were represented in our student interviewees; three were homogeneous groups (composed of all lower scorers), while 10 were heterogeneous groups (with a mix of high, mid-, and low scorers). Interviewee profiles with pseudonyms to maintain students' confidentiality are provided in Table 2. Emerged themes and quotes are summarized in Table 3.

Interview data were recorded, transcribed verbatim, and coded through multiple transcript readings by both authors. Specifically, we employed thematic analysis of the student interviews using ATLAS.ti v. 7.1 software. One researcher created a codebook related to how students perceived the group work and how it influenced their learning. Then the researchers worked together to refine the codebook and identify and characterize themes represented by the codes (Braun and Clarke, 2006).

We were interested in determining how statements made during interviews might be represented in comments on student evaluations of their peers, and we also believed these comments could provide insight into how students gave feedback to their group members to answer research question 3. Triangulating data from various sources provided a mechanism to verify and support the breadth and overall representation of our understanding of what students meant in their comments (Morse *et al.*, 2002). Students from 65 groups provided a total of 1341 entries on the midsemester and end-of-semester peer evaluations. We excluded 120 blank entries without text from further analysis. Of 26 students who failed to leave comments, seven were high-scoring students, nine were lower-scoring students, and 10 were midscorers. Most of the students (20) were from high-performance groups, and only three were from low-performance groups. We analyzed the remaining 1221 comments to provide insight into what students told their peers to help them improve group dynamics.

Because these comments could help determine whether or not students were taking advantage of this as a vehicle to express positive as well as negative perceptions, we began with categorizing a priori codes into three categories: positive, neutral, and negative indicators/experiences. We then added more a priori codes elicited from theoretical models of collaborative or cooperative learning that focused on explaining student motivation and engagement. For example, social interdependence theory (Johnson and Johnson, 2009) was considered, because the analysis aimed to examine how student perceptions of group work could derive from both interactions between individual students and group-level attributes that
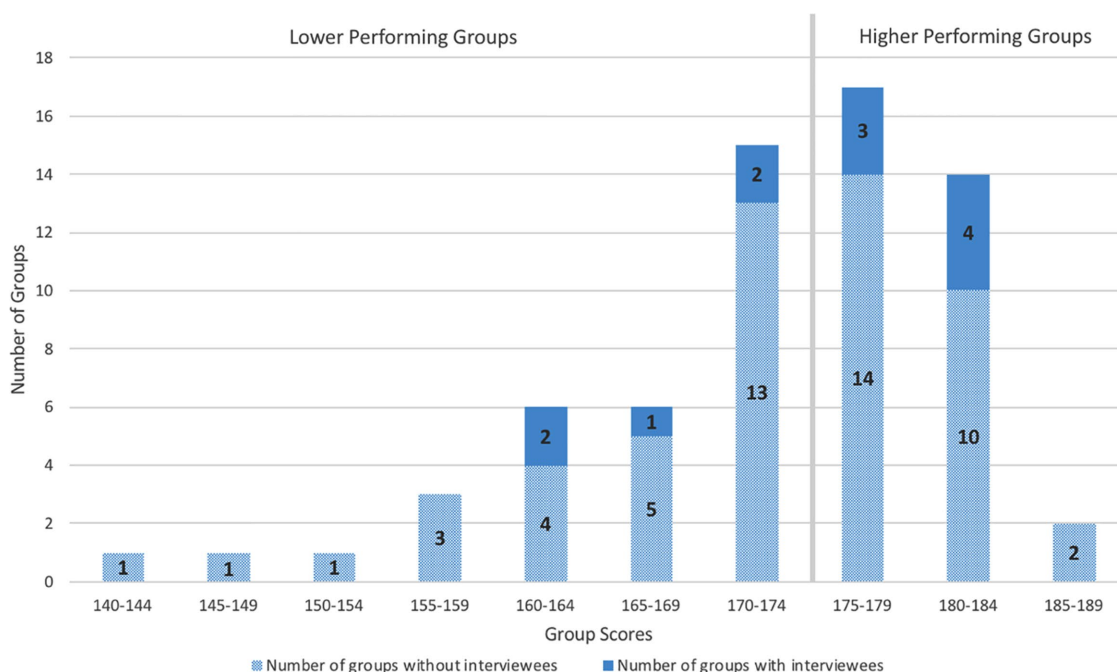
FIGURE 2. Performance levels of high- and low-performance groups. To select a balanced numbers of high- and low-performance groups, we divided groups into higher (> 175) and low-performance groups (< 175) using the median (175). Equal numbers of interviewees were selected (dark blue) from both types of groups.

support effective group work. A priori codes that reflected students' perceptions of group work in peer-evaluation comments and interviews are represented in Appendix C in the Supplemental Material. We initiated the coding of peer-evaluation comments by using these codes and added emerging codes as necessary. The first author (Y.C.) initiated the analysis and then the second author (P.B.) reviewed the analysis. Initially, 21 codes, including 11 codes for positive perspectives and 10 codes for negative perspectives on collaborative learning, were used to analyze peer-evaluation comments. After the second author (P.B.) reviewed the codes, the codebook was modified

and clarified through discussion. Two codes were merged in both the positive and negative perspectives, and three more codes were added. Finally, both authors reviewed the coding together, discussing and merging codes through consensus, and adding emergent codes as needed. Interrater reliability (Cohen, 1968) between the two authors was achieved at 0.85, considered as very good strength of agreement (Altman, 1991). Among all codes, 18 codes (10 positive, six negative, and two neutral) finally emerged from the analysis of peer-evaluation comments (Appendix C in the Supplemental Material).

TABLE 2. Interviewee profiles

	Pseudonym	Individual scoring H/L	Group performance HG/LG	Description
1	Monica	H	HG	In a group with a lower scorer, Amy, and two other midscorers
2	Min	H	HG	Created a purposeful group with Chen, another higher scorer, and two lower scorers
3	Chen	H	HG	In a group with Min and two lower scorers
4	Beth	H	HG	In a group with two lower scorers and another higher scorer
5	Ginger	H	HG	Higher scorer in a group with three midscorers
6	Chloe	H	LG	Higher scorer in a group with three midscorers
7	Nora	H	LG	Only higher scorer in a group with a midscorer and a lower scorer
8	Ethan	L	HG	Only lower scorer in a group with two midscorers
9	Jenn	L	HG	In a group with another lower scorer and two midscorers
10	Amy	L	HG	In a group with higher scorer, Monica, and two midscorers
11	Emma	L	LG	In a group with three other lower scorers
12	Ruth	L	LG	In a group with three other lower scorers
13	Karen	L	LG	In a group with three other lower scorers
14	Brian	L	LG	In a group with three mid scorers
15	Zoe	L	LG	In a group with two higher scorers and lower scorers

H = Higher scorer, L = Lower scorer, HG = Higher-performing group, LG = Lower-performing group.

TABLE 3. Excerpts per theme from in-depth interviews

Themes	Quotes	H/L	HG/LG
Learning gains from group activities; engagement, content knowledge acquisition, understanding concepts, and application of the acquired knowledge	"We all study differently and so if I thought of something in a certain way, but they didn't get it, I **could verbalize and explain it to them.**" (Monica)	H	HG
	"I'd be like, 'I don't know what's going on' and she'd be like, 'Oh, it's this and this,' like she would explain it to me. I asked her and so she **elaborated on … went into a little deeper context** with me and **tried to explain in it my terms in a way that the teacher couldn't**." (Ethan)	L	HG
	"I guess I feel like the tests in this class were **more like application, like applying what you've learned to a different situation.**"(Amy)	L	HG
	"We all had very **deep discussions about each question and why the answer was what it was**.… discussing it with my group members, we learned more just discussing the information with each other." (Ethan)	L	HG
	"Group work definitely helped and **engaged me more** with the learning." (Brian)	L	LG
	"I could ask 'What was your answer?,' 'How did you get that? I don't understand'… we always do that. '**This one is correct and you answered this one, it was wrong**.'" (Karen)	L	LG
Positive experiences with group members	"I let them (group members) know 'the activities will be on the test. You should understand it.' I took my notes and **shared it with other group members**. Then I **cheered them to respond to the practice quizzes**." (Ginger)	H	HG
	"To be able to talk to people about it and kind of work through it and figure it out. It was **like a nice little support system**. You didn't feel like you were alone in the class." (Beth)	H	HG
	"You **can ask your questions** to your group especially in a such a huge class." (Nora)	H	LG
	"I could **text them, they help. Email them, they help** … I mean, any time I don't understand because someone there is always going to be able to help you." (Karen)	L	LG
	"'I really don't understand this, I need help really bad'… I will text one of them and say, 'Can you explain this to me real fast?' **they usually respond really quickly**." (Karen)	L	LG
Contrasting perceptions about the purpose of the group activities	"It was more **real-life situations** which I found interesting because it was **more relatable** and like scientific theories that would apply to you … based on **how it would affect something in real life**." (Beth)	H	HG
	"**Why do that and spend an hour** and going through all of that stuff when I could just finish it in ten minutes." (Jenn)	L	HG
	"**Something doesn't seem very important and sound silly. I don't see any relationships with my real life.**" (Zoe)	L	LG
Attitudes and behaviors negatively influenced group activities	"I just think she didn't know how to really make her point without **overriding everybody else**." (Chloe)	H	LG
	"I know she might **want to do it herself** but that's not what we're supposed to do in a group… I couldn't contribute or did something more … **she never offers**."(Ruth)	L	LG
	"**We're going to get her to work, give her the opportunity to do stuff** and 'Hey, Anne, we're kind of confused on this, will you come over here and help us?' and so she came over and then we went through one part of the worksheet and then Tory and I kept saying stuff and she'd be like, **'Oh, yeah, that's what I was thinking, too.'**" (Amy)	L	HG
Individual group member assumed responsibility to complete group work	"[Other group members said,] 'Oh, I hope you guys know a lot for the group test because I don't really know anything' basically telling me that 'Oh, I hope you know a lot so I can write off your intelligence and get a good grade.'" (Nora)	H	LG
	"[My group members] just sent me all copy and paste. **They didn't do any research and follow any APA format**. But they simply sent me the link of the resources and said, 'here is the link.' So I have to go back to the website that they'd found and had to summarize them." (Min)	H	HG
	"They'd send me like not very good information and **I'd have to redo their whole thing.** If they sent me their portion of the things, I had to go through all of their information and put it together… once I actually put a document together and submitted it, they probably never even look at it. **They didn't help me in a way that could have helped them because they never really even looked at it**." (Ethan)	L	HG
	"It's more difficult to get all the members to contribute equally especially if someone is … doing all the work and then other people feel like they can relax, '**You're doing all the work and I can just sit back and get a hundred**' sort of like **the prisoner effect**." (Ruth)	L	LG
Groups did not collaborate but merged and submitted outside group work	"To one person and then that person usually puts into a document and then sends it back to us to review and then we all kind of tweak it and make sure it's right before that person submits in the Dropbox." (Ginger)	H	HG
	"After everyone does their own part they submit the document back onto Facebook, we attach it and so someone can continue from the other person. But **we usually don't edit it**.… **we don't have someone like edits or synthesizes**." (Emma)	L	LG

RESULTS

Interviews

Students in Both High- and Low-Performance Groups Valued the Social and Cognitive Support Provided by Groups. Seven interviewees (four from high-performance groups and three from low-performance groups) mentioned the social and cognitive benefits of group work. Among four students in higher-performance groups, two were higher scorers and the other two were lower scorers. Ruth, like other students (even those in lower-performance groups), exhibited positive perceptions about the benefits of group work, saying, "Our group got along a lot better with just interacting with each other. I've definitely learned better in a group because I had the opportunity to kind of answer some of their questions which helps me understand it more." Amy, a lower scorer in a higher-performance group, also described group work as "a nice little support system" that helped her to feel like she was "not alone in the class." Ethan, another lower scorer in a higher-performance group, mentioned that having group members helped him to understand course concepts, as they "elaborated on the concept and went into a little deeper context with me and tried to explain in it my terms in a way that the teacher couldn't because there is just so many people." Higher scorers in higher-performance groups also acknowledged benefits of group work, as it allowed them to "have friends in the class who can discuss things with typical lectures" (Beth) and "study together to prepare for the exam" (Chen). (See Table 3 for additional interview excerpts for each theme.)

Students Assigned Roles Depending on Circumstances and Ignored Group Contracts. We were interested in determining students' perceptions of the support strategies (e.g., role assignment and creating and revising group contracts) used to increase individual accountability and appropriate use of social skills. In general, students did not perceive the support strategies as beneficial. Regarding assigning or rotating roles, roles were naturally assigned "with respect and making it fair" (Amy) and rotated. One interviewee mentioned, "We always rotated who would actually compile all of the finished work and put it into a document and submit it to the class" (Jenn). None of the interviewees used role assignments consistently, citing a variety of reasons. Some groups felt they were unnecessary. "Our group didn't really do the role assignments, it was just kind of 'You get it this time, I'll get it next time'" (Jenn). Others reported grappling with disorganized or absent students. "If you're not communicating as well … you don't get to do the part that you want to do" (Emma). "Even though we assign each other the role, not everybody follows it in a way and they usually forget about it, it's not something they are used to having to do" (Min). Finally, some students expressed anxiety about assuming certain roles. "You were supposed to assign someone each of those jobs but I found it kind of limiting sometimes because … I got put as manager except I don't like raising my hand and talking in front of people and this other guy in my group did and he knew a lot of the answers" (Beth). Another interviewee commented that with such a large class they "never really had to use that role because we weren't really asked questions like in class about stuff" (Zoe).

In terms of the group contract, students described it as "another assignment that we had to do so we could leave" (Karen). Although some groups set punishment rules, such as "treat a coffee for all in the group if you missed a class without communication with group members" (Chen), group members reported that they rarely abided by the rules from their contracts. In fact, most of the groups (59 out of 65 groups) simply resubmitted their original contracts without revision at the midpoint of the semester, with only seven groups revising their contracts to provide better feedback to one another. One group revised their contract following discussion about lack of compliance with the initial contract. However, the group members still did not comply: "They all said, 'Sure' and then they don't do it" (Min). We also found peer pressure was significant when it came to opting not to follow through on punishments or criticize group member's behaviors. "I'd never tell these guys, because I didn't want them to say something about me" (Ethan).

We did find that students perceived the group contract as useful when it came to setting up communication methods. For example, "we all got our phone numbers and everything" (Min) and "when missing a class, text to all" (Beth). However, in well-functioning groups, the group contract was not needed. Six interviewees, five from high-performance groups and one from a low-performance group, mentioned that their groups did not need to use the group contract, because they encountered few challenges. For example, members from high-performance groups agreed that "we know we would follow through when we had stuff and we never really had to use the group contract" (Brian). As one of the interviewees mentioned, "I think if there had been problems in our group, the contract would've been a bigger role" (Nora).

Students in Both High- and Low-Performance Groups Reported Social Loafing. Regardless of group performance, most of the interviewees reported social loafing issues during group work that resulted in them having to assume responsibility for submitting work for others. Min, a higher-scoring student in a high-performance group, stated, "[My group members] just sent me all copy and paste. They didn't do any research and follow any APA format. But they simply sent me the link of the resources and said, 'here is the link.' So I have to go back to the website that they'd found and had to summarize them." Lower-scorer Ruth, who was in a low-performance group, commented that "it's more difficult to get all the members to contribute equally especially if someone is … doing all the work and then other people feel like they can relax. You're doing all the work and I can just sit back and get a hundred sort of like the prisoner effect."

Student remarks differed when they described the same situation, one in which they had to take over and perform work to make up for noncontributing students. Min, a high-scoring student in a high-performance group, criticized other group members' inadequate participation, even though her group achieved higher scores on group work. "Normally they don't do the pre-quiz because it's not for a grade. But like I do it just for my good, but they don't. It feels like they don't care and know I'll do it anyway" (Min). While Ruth, a lower scorer in a lower-performance group described her experience as "getting sucked into all of the group work." Hard-working students reported experiencing the "sucker effect," in which they began to pull back on the amount of work they contributed to force noncontributing members to work: "We were like 'We'll say we don't understand

this part so she can do it' and then she like didn't know and just held it until the end of the class and was like, 'I don't get it'… and so my friend got kind of bothered and she snatched it from her and did it" (Monica).

Perceptions of the Value of Group Activities Differed between High- and Low-Scoring Students. We found that three of the seven higher-scoring students interviewed, two from high-performance groups and one from a low-performance group, remarked that they found group activities were critical for their learning and that they provided an opportunity to apply what they learned from lecture to their real life. For example, Beth, who was a higher scorer, reported that group work was "more real-life situations which I found interesting because it was more relatable and like scientific theories that would apply to you … based on how it would affect something in real life." Six of the seven lower scorers interviewed, in both high- and low-performance groups, tended to perceive group activities as time-consuming. For example, Jenn asked, "Why do that and spend an hour going through all of that stuff when I could just finish it in ten minutes?" She felt, "It was like a stumbling around so that didn't really help with the learning." Zoe, another lower scorer, echoed this sentiment, describing group work as "something [that] doesn't seem very important and sound[s] silly," and said, "I don't see any relationship with my real life."

High- and Low-Performance Groups Left Very Similar Comments on Peer Evaluations. We compared patterns observed between the overall frequency of peer-evaluation comments (positive or negative, as well as those expressing specific ideas) between students in high- and low-performance groups. We calculated the frequency of the code by total number of submitted comments. We found that there was no statistically significant association between group performance levels and the trends on the positive or negative comments; $x^2(34) = 30.78$, $p = 0.626$. In addition, regardless of group performance levels, three categories of ideas were commonly expressed (appearing in more than 10% of all comments) in both groups in both the first and second peer-evaluation surveys (Figure 3). A fourth and fifth category did differ in its frequency between groups of different performance levels, which we will describe later. No distinctive patterns were found in the comments categorized as no responses or not applicable.

The most common idea mentioned (>55% of all codes present) in peer evaluations involved individual accountability. Individual accountability involves students completing their own work (personal accountability) and facilitating other students completing their work (accountability to the group; Johnson *et al.*, 2014). Students mentioned aspects that reflected positively on their peers: "always doing her part," "completing assigned works in a timely and efficient manner," and "pulls his weight and helps with all group work." See Appendix C in the Supplemental Material for additional codes, summaries of ideas represented in those codes, and example quotes. We found that comments about individual accountability were equally predominant in high- and low-performance groups and at the midpoint and endpoint of the semester ($\chi^2 = 4$, $p = 0.261$).

The second most prevalent idea that emerged in peer-evaluation comments concerned the cognitive learning supports provided by peers (22–31% of comments contained this idea).

Students seemed to appreciate aspects of the promotive interactions mentioned by Johnson *et al.* (2014) that enhanced social constructivism, such as engaging in a dialogue with their peers to ask and answer questions, share reasoning, and build upon each other's understanding until they reach mutual agreement (Phelps and Damon, 1989; Slavin, 1991). Student evaluation comments mentioned students "explaining difficult concepts" or a peer "helps us understand" or "takes the time to learn the material so she can teach others." The degree to which students mentioned learning supports appeared to differ between higher- and low-performance groups when comparing the midpoint to endpoint survey, but these differences were not significant ($\chi^2 = 3$, $p = 0.223$).

The third most frequent idea that emerged in the peer-evaluation comments concerned students providing procedural support. This idea was present in more than 17% of all comments. However, there was no statistical difference between high- and low-performance groups mentioning procedural support ($\chi^2 = 4$, $p = 0.261$). Procedural support involved helping to complete steps to finalize a task that were not connected to cognitive activities but aided in making decisions and completing assignments in a timely manner. Examples of comments that were coded for procedural support mentioned their peers helping to "keep everyone focused and turn in our papers at the end of class" or being "very organized and keeps the group on track, schedules group meetings."

The fourth most frequent idea that was mentioned differed depending on the time of the semester the evaluation was given and the performance level of the group. At the midpoint of the semester, the fourth most frequent idea mentioned in high-performance groups involved social/interpersonal communication skills (10.3% of comments). This idea included the ability to get along with group members and communicate effectively, for example, "displays a positive attitude and is very encouraging," "easy to talk to and easy to get along," and "very flexible and open to everyone's contributions and a fun person to have in the group." Social perceptiveness has emerged as a major factor in predicting group performance (Woolley *et al.*, 2015). Interestingly, social/interpersonal communication was not a frequent idea mentioned in low-performance groups until the end of the semester. In those groups, the fourth most frequent type of comment left at the midpoint was to not leave any comment at all (13% of comments left in low-performance groups were left intentionally blank). By the end of the semester, social/interpersonal skills rose in its frequency to become the fourth most frequent code mentioned in evaluations from students in low-performance groups (10.3% in the end-of-semester peer evaluations).

Students in high- and low-performance groups also differed in the degree to which they mentioned that their peers provided positive interdependence, with promoting the group's success by working together seen as a fifth frequent idea. Calculating the percentage of frequency of the positive interdependence code by the total number of submitted comments, students in high-performance groups mentioned positive interdependence in 12% of all comments on the final peer evaluation, up from only 6% on the midsemester evaluation. Lower-performance groups mentioned positive interdependence on the midsemester evaluations (7% of comments); however, this level dropped to less than 2% of the final evaluation.
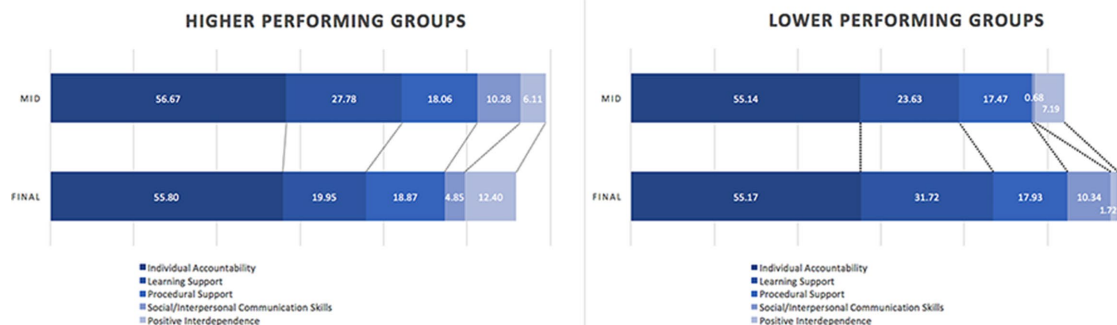
FIGURE 3. Frequency comparison for the top four ideas provided as anonymous comments on peer evaluations. Student groups were categorized as higher-performing ($n = 32$) and lower-performing ($n = 33$) based on the rank-ordered group assignment scores and averaged group test scores. A total of 1221 comments on peer evaluations were coded using 21 a priori categories. Each number (%) in the figure is calculated as the frequency of the code/total numbers of submitted comments.

When comparing comments between heterogeneous groups of either higher- or lower-performance levels, we found that students in heterogeneous groups leave more comments with a negative perspective (31.25% on midpoint evaluation and 28.57% on final evaluation) compared with homogeneous groups (nothing on mid and 18.18% on final), regardless of the timing of the evaluation.

Influence of Group Work on Individual Learning Achievement within Groups. We compared the mean score differences between group and individual test scores in units 1 and 4 to understand the influence of group work on individual learning. The random intercepts model results show that group performance levels influenced score differences between individual and group tests from unit 1 and unit 4. As shown in Table 4, both high- and low-performance groups actually decreased the mean score differences between group and individual test scores from unit 1, $F(1, 122) = 19.83$, $p < 0.005$, to unit 4, $F(1, 100.72) = 20.29$, $p < 0.005$. In unit 1, the main effect of group performance levels on mean score differences between individual and group test scores are significant ($p < 0.005$), estimating that the mean score differences of high-performance groups are 12.92 (SE = 2.68) higher than those of low-performance groups. In unit 4, the estimated mean score differences of high-performance groups are 11.85 (SE = 2.9) higher than those of low-performance groups.

Within groups, high-performance group score differences decreased significantly from unit 1 (M = −12.04, SE = 1.83) to unit 4 (M = −6.78, SE = 1.70), $p < 0.005$. Low-performance groups also decreased from unit 1 (M = 24.96, SE = 2.26) to unit 4 (M = 18.63, SE = 2.10). However, high-performance groups demonstrated a greater reduction (43.7%) in the learning gap between group members compared with low-performance groups (25.3%).

Peer-Evaluation Ratings: Indicative of Performance but Possibly Biased

We found that, when students were asked to provide quantitative ratings of their peers, the ratings differed based on their groups' level of performance, $F(1, 164) = 12.97$, $p < 0.001$. In groups ranked as higher performing, students gave higher mean peer ratings to group members (M = 102.63, SD = 5.83) than students in low-performance groups (M = 92.32, SD = 19.15). As shown in Figure 4, in terms of individual differences between group members, t tests revealed statistically significant differences, $t(166) = −20.21$, $p < 0.001$, $r = 0.32$. Students who earned lower test scores received higher mean peer-evaluation ratings when they were in high-performance groups (M = 101.21, SD = 1.99) rather than in low-performance groups (M = 86.96, SD = 25.63). Higher scorers received relatively similar peer-evaluation ratings from both high-performance groups (M = 101.78, SD = 5.96) and low-performance groups (M = 97.35, SD = 7.04). This resulted in greater mean score differences between higher and lower scorers in low-performance groups (mean score differences = 14.25) than between higher and lower scorers in high-performance groups (mean score differences = 5.53). In other words, students in low-performance groups were holding lower-scoring students to a greater degree of accountability on peer ratings than lower scorers in high-performance groups.

TABLE 4. Estimates of fixed effects of score gap between group and individual scores in units 1 and 4

		Estimate	SE	df	t	Sig.	95% Confidence interval	
							Lower bound	Upper bound
Unit 1	Intercept	−24.96	2.26	122	−11.06	0.000	−24.42	−20.49
	High-performance groups	12.92	2.90	122	4.45	0.001	7.18	18.67
	Low-performance groups	0	0	—	—	—		
Unit 4	Intercept	−18.63	2.10	69.57	−8.89	0.000	−22.81	−14.45
	High-performance groups	11.85	2.63	100.72	4.5	0.001	6.63	17.08
	Low-performance groups	0	0	—	—	—		
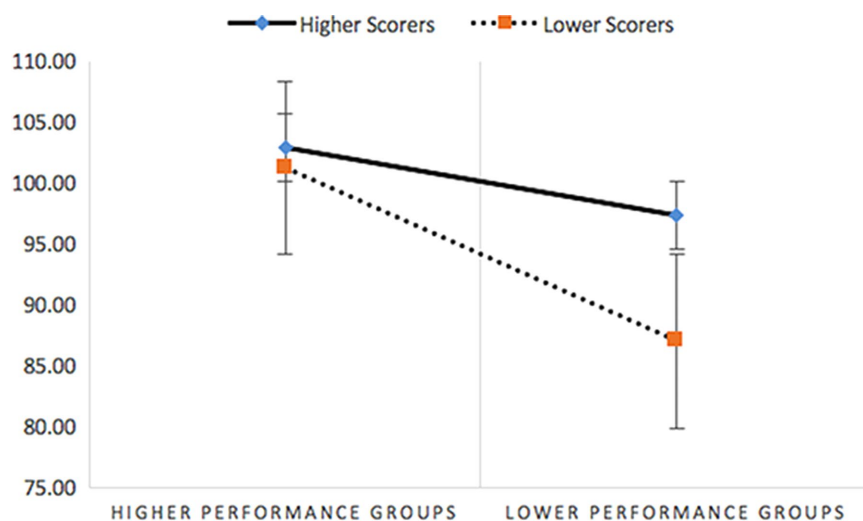
FIGURE 4. Peer-evaluation ratings and group performance. The average numerical ratings (out of 100) that students wrote on peer evaluations differed for high- and low-performance groups. The average rating for all students in high-performance groups (M = 102.63, SD = 5.83) was higher than students in low-performance groups (M = 92.32, SD = 19.15). Lower-scoring students received higher mean peer-evaluation ratings when they were in high-performance groups (M = 101.21, SD = 1.99) than in low-performance groups (M = 86.96, SD = 25.63). Higher scorers received relatively similar peer-evaluation ratings from both high-performance groups (M = 101.78, SD = 5.96) and low-performance groups (M = 97.35, SD = 7.04), t (166) = −20.21, p < 0.001, r = 0.32).

DISCUSSION

To understand group work in a real-world classroom setting, we conducted our study in a classroom that allowed students to self-assemble into groups. We conducted a mixed-methods analysis to investigate how students' personal scoring in the course and their groups' levels of performance on group tasks affected their perception of group activities. Much of a student's individual performance was calculated after group work occurred. Group and individual performance, though, may not be completely independent of each other, because students who score highly on individual tests could have an increased effect on their groups' scores on assignments and tests, and groups that worked well together could have had a positive effect on individual grades. Collection of an initial independent measure of a student's abilities (e.g., grade point average [GPA] or pre-test scores) or analysis of discourse practices during group work would be needed to further elucidate the relationship between individual ability on group performance. We also interviewed students on their use of strategies such as role assignment and group contracts, and we compared the comments and ratings from anonymous peer evaluations to determine what differences could be observed between high- and low-performance groups. We found that, regardless of group performance levels, salient elements that might affect students' perceptions of the group work emerged. We did find differences in perceptions of group work, some comments on peer evaluations, and ratings between individuals. Our hope was to uncover factors that could explain why groups might not be working effectively and to provide faculty with a better mechanism to identify and solve group problems. However, we did not conduct empirical tests for these observations, so we will use our results to suggest future research to expand on our findings.

Without Supervision, Students Fail to Use Role Assignment and Group Contracts to Their Fullest

In our study, students were asked to self-assign to specific roles and rotate the roles within the group activities in a very large classroom without supervision (either from undergraduate peers or graduate students). It has been reported that structured discussion with role assignment enhances students' engagement in interactive information sharing (Mesmer-Magnus and DeChurch, 2009), knowledge construction, knowledge transfer (Kane et al., 2002), and equitable participation (Savadori et al., 2001). In the in-depth interviews, however, students did not report engaging in what Chan (2001) described as detailed "problem-centered" discourse that involved recognition of the problem, formulation of questions, and construction of explanations. Instead, they reported that they felt that they did not need official roles. Some students struggled assigning roles to absent or inactive members, and other students commented that certain roles were never used because of the number of groups and the difficulty hearing ideas from all groups during class discussion. Instead, they reported using roles to subdivide labor, merging individual contributions to a final document without critically evaluating members' submissions, failing to communicate, and completing a task at the very last minute so that it could not be reviewed by all group members (see comments in *Students in Both High- and Low-Performance Groups Reported Social Loafing*). All these obstructions to the collaboration process erode the sense of trust needed for the social interactions required for social interdependence. Students in our groups were only provided with brief descriptions of what group roles entailed and were not explicitly trained to use roles effectively. Several research groups have attempted to test the effect of providing scripts to enhance students' use of cognitive prompts during group work (O'Donnell, 1996; Gillies, 2003; Brewer and Klein, 2006). It would be interesting to determine whether these scripts can be useful in the very large lecture settings we encountered.

Group contracts provide a mechanism to initiate discussion of expectations and reservations, to strengthen social skills, and to build interpersonal relationships critical to effective group work (Oakley et al., 2004; Davies, 2009; Shimazoe and Aldrich, 2010). It is clear from our interviews, however, that students did not view the completion of a group contract as a significant vehicle to strengthening interdependence or other key social skills. They felt the assignment was cursory, just something to be completed without depth of thought. Student teams become progressively more collaborative and productive over time (Hong et al., 2014), with successful teams demonstrating constant levels of socioemotional and procedural support (Kwon et al., 2014). So, rather than assigning a simple contract at the beginning of the course, instructors may need to provide more opportunities for deeper interactions that

build trust and a sense of belonging (Kreijns *et al.*, 2003), to be conscious of the fact that trust can be easily eroded when members lower their individual participation and commitment toward quality work (Kreijns *et al.*, 2003), and to provide greater opportunities for identifying and resolving conflict (Brooks and Ammons, 2003). We found that students were not relying on group contracts to set punishments for lack of participation and that other mechanisms need to be employed to help groups resolve conflicts (see *Peer Ratings May Be Biased against Low-Scoring Students in Low-Performance Groups*). An alternative assignment with more utility might be to engage in an initial team-building activity in which students share contact information to begin to build socioemotional interactions like getting to know one another by sharing hobbies, interests, and experiences (Oakley *et al.*, 2004). Researchers could test the efficacy of such an assignment using measures like the Team Interdependence (Van der Vegt *et al.*, 2001), Team Cohesiveness (Carless and De Paola, 2000), or Psychological Safety (Edmondson, 1999) surveys.

Peer-Evaluation Comments Are Not Useful for Identifying Group Dysfunction

Anonymous peer evaluations have been suggested as a method for identifying inequity and other problematic group issues (Wenzel, 2007; Aggarwal and O'Brien, 2008). Peer evaluations have been shown to reduce the incidence of free riding and to improve student attitudes toward groups and group projects if they are done early and frequently (Feichtner and Davis, 1984; Brooks and Ammons, 2003). Using performance on group assessments to identify high- and low-performance groups, we asked whether performance could be differentiated using responses to students' peer-evaluation comments and numerical ratings. We also used students' responses to in-depth interviews to provide context and explanations for our observations. Several assumptions may limit our findings: We must assume that students were willing to communicate honestly about problems that were occurring during group work. We also must rely on students' ability to recognize what constitutes effective group functioning. Also, we assume that, because a group is performing well on assignments, they are functioning better than a group that is not earning high scores on group assignments and tests. It is possible that high-performance groups do better on assignments because one high-scoring student has taken on more of the responsibility. However, it is clear from our analysis of how the score gap between individual and group tests is minimized in high-performance groups (Table 4) that learning as measured by the level of scoring on individual tests improves for low-scoring members of these groups. This study does not clarify the reason for this improvement, but high-performance groups had a higher percentage of high-scoring students, so access to these students could be one important variable. With this in mind, we were surprised that comments on anonymous peer evaluations do not adequately distinguish between high- and low-performance groups, considering that the major comment codes were similar in both these groups at the middle and end of the semester, with only a few subtle differences. The data did, however, spur us to hypothesize that one of the major differences between high- and low-performance groups may derive from student attitudes and appreciation of the value of group work to their learning overall.

We found that all groups, regardless of their success, clearly viewed individual accountability as their basic responsibility as a group member. This constituted the most frequent comment in peer evaluations from both high- and low-performance groups at both the midpoint and end of the semester. Individual accountability, although critical and mentioned as the primary comment in peer-evaluation comments, should be seen as a minimal requirement for effective group work. In well-designed group work, task assessments and grading for individual contribution play a greater role in promoting and enforcing individual accountability than self-regulation from group members. From a cognitive learning perspective, consensus building and coconstruction of knowledge (Solomon, 1987; Latour and Woolgar, 2013) constitute a more effective method for enhancing learning during group work than promoting completion of tasks by divvying up the work. Students bring misconceptions about their fellow students' attitudes and abilities to group work that prevent them from distinguishing peers who were deliberately failing to contribute from peers who were struggling academically and contributing to the best of their abilities (Freeman and Greenacre, 2011). Without some clarity, peer evaluations can reinforce these attitudes. Providing students with a more careful evaluation of cognitive behaviors (e.g., asking to list individual efforts toward completing group activities such as editing, writing portions, finding references) and creating task structures that ensure that students provide explanations or elaborations could help produce more turn-taking, productive discourse, and appreciation that students bring varying levels of expertise and contributions to understanding. Student groups that exhibit networks of communication with frequent interpersonal interactions exhibit higher cognitive complexity and performance overall (Curşeu *et al.*, 2012).

It is interesting that peer-evaluation comments related to communicating effectively were less frequently observed in lower-performance groups. Students in lower-performance groups do eventually mention this, but not until later in the semester. In addition to communication, students in high- and low-performance groups expressed an appreciation for the value and importance of positive interdependence in comments on midsemester evaluations. However, students in low-performance groups were less likely to mention positive interdependence in the end-of-semester evaluations. Reliance on group members to achieve common goals has been consistently identified as a critical factor in increasing collaboration, and its effect on achievement (Slavin, 1991; Webb and Palincsar, 1996; Johnson *et al.*, 2007; Scager *et al.*, 2016). High-performance groups may work better because they perceive the value of group work from the beginning of the course and recognize their group members for providing features of learning support, communication, and positive interdependence (Kwon *et al.*, 2014). The absence of comments about positive interdependence in low-performance groups at the end-of-semester evaluations is not unexpected: they were not performing effectively on assignments. However, in considering the implications for instructors, these are not the comments that are showing up most frequently, so it would be hard to identify dysfunctional groups using this feature.

There are several limitations to our findings, and additional research is needed to corroborate the correlations we observed between these self-report measures and academic performance. Our finding that lower-scoring students are critical of group

work comes from interview data. It would be interesting to extend our work by surveying an entire course with a group-work satisfaction survey to determine whether our findings extend to all students within the course. We did not ask students to provide detailed self-reports on their own contributions to each assignment. These may prove useful for further studies that could compare individual self-reports (self-evaluations) with anonymous peer evaluations to correlate with individual and group performance measures. Interview and reflection journals may also prove important in determining the extent to which self-report indicators can be employed with either individual or consensus self-assessment measures. Instructors could then provide adaptive scaffolding to support students' monitoring and reflection on their learning depending on the progress of group work. For example, when groups perform poorly, instructors would be able to provide extra procedural support (e.g., providing a checklist, suggesting using track-changes options to record individual contributions) or strategic support (e.g., providing summary notes or extra review sessions).

In addition to their ineffectiveness at identifying group dysfunction, peer evaluations may negatively impact how students interact with one another. During the interviews, students reported that they were concerned about the possibility of their identities being revealed in their peer-evaluation comments and that this would negatively affect their group relationships. It is obvious that students experienced peer pressure even with an online peer-evaluation system designed with a confidential log-in system to ensure anonymity.

As Bacon et al.'s (1999) study indicated, use of evaluations can be negatively associated with good team experiences. Strong and Anderson (1990) observed that students preferred other factors—including group cohesiveness, small team size, the option to divorce a team member, or the option to leave a team—as stronger countermeasures for reducing free riding.

Peer Ratings May Be Biased against Low-Scoring Students in Low-Performance Groups

We also examined the numerical ratings that students provided for one another as a summative assessment of their effort on group work throughout the semester. This reward or punishment system is designed to deter social loafing, to motivate individual students to enhance their performances, and to account for inequities in students' contributions. Students' most common complaints about group work involve uneven contributions from group members (Livingstone and Lynch, 2000; Aggarwal and O'Brien, 2008; Pauli et al., 2008; Shimazoe and Aldrich, 2010; Hall and Buzwell, 2012). However, there are several problems with using peer ratings to evaluate contributions. First, students have expressed discomfort using ratings that they view as criticizing their friends (Williams, 1992). Student raters may not have the ability to distinguish between high and low contributions, and their ratings may be influenced by their own experiences and norms within their groups (Loughry et al., 2007). Underperforming students tend to over- or underrate their individual progress, need for support, and understanding (Winne and Jamieson-Noel, 2002). Students often resort to giving all students equal marks on a holistic evaluation compared with a categorical evaluation that queries about individual skills (Lejk and Wyvill, 2001), and unadjusted ratings that students receive from their peers do not correlate with

course grades (Zhang and Ohland, 2009). Finally, students within a group may intentionally inflate or reduce the contribution of members due to friendship (Zhang et al., 2008).

We found that the ratings students gave their peers did correlate with group performance: high-performance groups gave higher ratings to their members compared with low-performance groups. So, groups were clearly using ratings in a small way to indicate poor performance, but their use may only be accurate for high-scoring students who received high ratings in both higher- and low-performance groups. Ratings may not provide an unbiased measure of individual effort or accountability for lower-scoring students, who were more likely to receive lower peer ratings in low-performance groups relative to lower-scoring students in high-performance groups. In the successful student groups that Scager and colleagues (2016) interviewed, students expressed a sense of empathy for their peers who contributed less, noting that it may have been "beyond (their) capabilities at that point." This may echo what we have seen in our analysis of peer ratings. High-scoring students in high-performance groups may recognize that their lower-scoring peers are simply contributing to the level of their ability. It was clear from their peer-evaluation comments that students in high-performance groups value communication and the sense of communal effort that is afforded by positive interdependence to a greater degree than students in low-performance groups. It is also possible that lower-scoring students in low-performance groups actually participate less than those in high-performance groups and actually earn those lower scores. Considering that students in both high- and low-performance groups complained of social loafing and free riding in interviews and that there are discrepancies within ratings for low-scoring students, there are clearly problems with using peer ratings to evaluate contributions to group work.

Alternative measures may be required to ensure equitable assignment of group contributions to account for biases in peer ratings. We used the expected contributions adjustment to modify group scores based on peer ratings. However, Zhang and Ohland (2009) found this to have higher absolute error and low correlation with true contribution to the team as measured by Monte Carlo simulations or actual class data. They recommend using a between-group difference adjustment that compares students to other students in matched performance groups, explaining that "as peer and self ratings are related to the quality of group work, ratings from groups with different group scores are not directly comparable" (Zhang and Ohland, 2009, p. 295). This is only a partial solution to the problems mentioned earlier. It may also be useful to ask students to specify the cognitive and functional efforts contributed by each group member on each assignment (e.g., an acknowledgments section in which they select from a list of tasks, including conducting a literature search, writing, creating figures, and editing). This could be very effective if students are also asked to reflect qualitatively on the social support (motivation, response to criticism, adaptability, creativity, and attitude) provided by members of their group.

For instructors with extremely limited time, there are also several scales specifically developed to identify conflict (Jehn and Mannix, 2001) or satisfaction (Van der Vegt et al., 2001) in student teams. These may be a quick way to identify problems in the early stages of group work. Organizations can provide resources for managing team conflict (Manktelow et al., 2017), and there is evidence that conflict management can improve

team performance, even in student teams (Tekleab *et al.*, 2009). Assigning group members descriptions of group dysfunction and asking students to identify strategies for solving the problem as a way to mediate their own conflicts (Lerner, 1995) or training students in reflexivity (O'Neill *et al.*, 2017) could provide mechanisms for groups to resolve conflicts before they result in dysfunction.

Are Time-Consuming Strategies That Facilitate Group Work Worth All the Effort?

In this study, we discovered that students of all abilities valued group work for various reasons in addition to how it benefited their learning. In student interviews, we observed that only about one out of seven lower-scoring students perceived group work and instructional facilitating activities to be beneficial for their learning performance (compared with three of seven high-scoring students). Lower-scoring students commented that they found group work to be beneficial primarily as a *comfort zone* or *the incentive* for them to go to class because they have friends to talk with. Motivation and social cognition are important mediators of group work (Slavin, 2014). Researchers have suggested that the quality of interactions between group members may be more predictive of learning gains than ability grouping and have recommended that an individual student's attitude, motivation, or personality traits may provide a better predictor of group success than cognitive ability alone (Webb *et al.*, 2002; Woolley *et al.*, 2010). In laboratory simulations of group work in which students are tasked with solving visual puzzles, brainstorming, making collective moral judgments, and negotiating limited resources, Woolley and colleagues (2010) have identified a single latent factor they call general collective intelligence—supported by strong interitem correlation on different tasks—that strongly predicts the groups' ability to solve tasks. Collective intelligence appears to depend both on the composition of the group (e.g., average member intelligence and, more importantly, social sensitivity) and how group members interact when they are assembled (e.g., their conversational turn-taking behavior). Groups in which a few people dominated the conversation were less collectively intelligent than those with a more equal distribution of conversational turn-taking (Woolley *et al.*, 2010). This was substantiated recently in a study that demonstrated that students who indicate higher levels of comfort with their group members achieve higher learning gains, whereas lower learning gains occur in groups with a reported conversation dominator (Theobald *et al.*, 2017). Because we allowed students to self-select in this study, it is possible that purposeful group composition may yield different results, as peer relationship can be used as a predictor of student performance within a group (Klein and Mulvey, 1995; Chung *et al.*, 2018). It would be interesting to examine the role of social sensitivity, beliefs about learning, friendship, and social skills such as conversational turn-taking as predictors of group performance in the college science, technology, engineering, and mathematics classroom. Also, as one of the goals of this study is to understand the effectiveness of group work in an uncontrolled classroom, we did not measure students' prior knowledge levels (e.g., Scholastic Aptitude Test scores, college GPA, pretest). To further investigate the influence of the specific group-based activity, we recommend administering pre- and posttests.

We found that, regardless of group composition or group performance level, students were likely to report positive experiences with group members. When group members function interdependently, collective efficacy beliefs have been shown to provide a greater impact on performance: groups with higher self-efficacy beliefs were more likely to encourage group members to use resources more effectively (Bandura, 2001) and to engage in higher-quality discussions (Wang and Lin, 2007). Thus, the establishment of collective group efficacy may be well worth the effort to promote group-learning performance (Gully *et al.*, 2002).

REFERENCES

Aggarwal, P., & O'Brien, C. L. (2008). Social loafing on group projects: Structural antecedents and effect on student satisfaction. *Journal of Marketing Education*, *30*(3), 255–264.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

American Association for the Advancement of Science. (2010). *Vision and change: A call to action*. Washington, DC. Retrieved August 24, 2018, from www.visionandchange.org/VC_report.pdf]

Armstrong, N., Chang, S.-M., & Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. *CBE—Life Sciences Education*, *6*(2), 163–171.

Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, *23*(5), 467–488.

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.

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1–26.

Barkley, E. F., Cross, K. P., & Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty*. Hoboken, NJ: Wiley.

Barr, D. A., Gonzalez, M. E., & Wanat, S. F. (2008). The leaky pipeline: Factors associated with early decline in interest in premedical studies among underrepresented minority undergraduate students. *Academic Medicine*, *83*(5), 503–511.

Bartlett, R. L. (1995). A flip of the coin—a roll of the die: An answer to the free-rider problem in economic instruction. *Journal of Economic Education*, *26*(2), 131–139.

Batz, Z., Olsen, B. J., Dumont, J., Dastoor, F., & Smith, M. K. (2015). Helping struggling students in introductory biology: A peer-tutoring approach that improves performance, perception, and retention. *CBE—Life Sciences Education*, *14*(2), ar16. doi: 10.1187/cbe.14-08-0120

Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., … Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-Based Reform of University Physics*, *1*(1), 2–39.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Brewer, S., & Klein, J. D. (2006). Type of positive interdependence and affiliation motive in an asynchronous, collaborative learning environment. *Educational Technology Research and Development*, *54*(4), 331–354.

Brickman, P., Gormally, C., Francom, G., Jardelezz, S. E., Schutte, V. G. W., Jordan, C., & Kanizay, L. (2012). Media-savvy scientific literacy: Developing

critical evaluation skills by investigating scientific claims. *American Biology Teacher*, *74*(6), 374–379. doi: 10.1525/Abt.2012.74.6.4

Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. *Journal of Education for Business*, *78*(5), 268–272.

Brown, P. J. (2010). Process-oriented guided-inquiry learning in an introductory anatomy and physiology course with a diverse student population. *Advances in Physiology Education*, *34*(3), 150–155.

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

Chan, C. K. K. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science*, *29*(6), 443–479. doi:10.1023/A:1012099909179

Chang, M. J., Eagan, M. K., Lin, M. H., & Hurtado, S. (2011). Considering the impact of racial stigmas and science identity: Persistence among biomedical and behavioral science aspirants. *Journal of Higher Education*, *82*(5), 564–596.

Chang, Y., & Hannafin, M. J. (2015). The uses (and misuses) of collaborative distance education technologies. *Quarterly Review of Distance Education*, *16*(2), 77–92.

Chapman, K. J., & Van Auken, S. (2001). Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education*, *23*(2), 117–127.

Chung, S., Lount, R. B., Jr., Park, H. M., & Park, E. S. (2018). Friends with performance benefits: A meta-analysis on the relationship between friendship and group performance. *Personality and Social Psychology Bulletin*, *44*(1), 63–79.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.

Couch, B., Brown, T., Schelpat, T. J., Graham, M. J., & Knight, J. K. (2015). Scientific teaching: Defining a taxonomy of observable practices. *CBE—Life Sciences Education*, *14*(1), ar9.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods research*. Los Angeles: Sage.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977.

Curşeu, P. L., Janssen, S. E., & Raab, J. (2012). Connecting the dots: Social network structure, conflict, and group cognitive complexity. *Higher Education*, *63*(5), 621–629.

Davies, W. M. (2009). Groupwork as a form of assessment: Common problems and recommended solutions. *Higher Education*, *58*(4), 563–584.

Deutsch, M. (1949). A theory of cooperation and competition. *Human Relations*, *2*, 129–152.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383.

Esterberg, K. G. (2002). *Qualitative methods in social research*. Boston: McGraw-Hill.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322.

Feichtner, S. B., & Davis, E. A. (1984). Why some groups fail: A survey of students' experiences with learning groups. *Journal of Management Education*, *9*(4), 58–73.

Freeman, L., & Greenacre, L. (2011). An examination of socially destructive behaviors in group work. *Journal of Marketing Education*, *33*(1), 5–17.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, *111*(23), 8410–8415. doi: 10.1073/pnas.1319030111

Gillies, R. M. (2003). Structuring Cooperative Group Work in Classrooms. *International Journal of Educational Research*, *39*(1–2), 35–49.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology*, *87*(5), 819–832. doi: 10.1037/0021-9010.87.5.819

Hall, D., & Buzwell, S. (2012). The problem of free-riding in group projects: Looking beyond social loafing as reason for non-contribution. *Active Learning in Higher Education*, *14*, 37–49.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... & Wood, W. B. (2004). Scientific teaching. *Science*, *304*(5670), 521–522.

Harkins, S. G., & Jackson, J. M. (1985). The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin*, *11*(4), 457–465. doi: 10.1177/0146167285114011

Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, *60*(7), 637–644.

Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology*, *105*(1), 89–107. doi: 10.1037/A0029691

Hong, H. Y., Chang, Y. H., & Chai, C. S. (2014). Fostering a collaborative and creative climate in a college class through idea-centered knowledge-building. *Instructional Science*, *42*(3), 389–407. doi: 10.1007/s11251-013-9289-y

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.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379.

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*(1), 15–29.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change*, *30*(4), 26–35.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). Constructive controversy: The educative power of intellectual conflict. *Change*, *32*(1), 28–37.

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 University Teaching*, *25*(4), 1–26.

Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education*, *29*(6), 751–768.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. doi: 10.1037/a0028347

Kane, A., Argote, L., & Levine, J. M. (2002). Social identity and knowledge transfer between groups. In *Annual Meeting of the Academy of Management, Denver, CO*.

Kaufman, D. B., Felder, R. M., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education*, *89*(2), 133–140.

Kilic, G. B., & Cakan, M. (2006). The analysis of the impact of individual weighting factor on individual scores. *Assessment & Evaluation in Higher Education*, *31*(6), 639–654.

Klein, H. J., & Mulvey, P. W. (1995). Two investigations of the relationships among group goals, goal commitment, cohesion, and performance. *Organizational Behavior and Human Decision Processes*, *61*(1), 44–53.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353. doi: Pli S0747-5632(02)00057-2 Doi 10.1016/S0747-5632(02)00057-2

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Kwon, K., Liu, Y.-H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education*, *78*, 185–200.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work—causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*(6), 822–832. doi: 10.1037//0022-3514.37.6.822

Latour, B. M. D., & Woolgar, J. (2013). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 26(1), 61–72.

Lejk, M., Wyvill, M., & Farrow, S. (1996). A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education*, 21(3), 267–280.

Lerner, L. D. (1995). Making student groups work. *Journal of Management Education*, 19(1), 123–125.

Livingstone, D., & Lynch, K. (2000). Group project work and student-centred active learning: Two different experiences. *Studies in Higher Education*, 25(3), 325–345.

Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, 67(3), 505–524.

Manktelow, J., Jackson, K., Edwards, S., Pearcey, E., Eyre, E., Cook, L., … Moss, I. (2017). Resolving team conflict: Building stronger teams by facing your differences. MindTools. Retrieved August 24, 2018, from www.mindtools.com/pages/article/newTMM_79.htm

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94(2), 535.

Michaelsen, L. K., Davidson, N., & Major, C. H. (2014). Team-based learning practices and principles in comparison with cooperative learning and problem-based learning. *Journal on Excellence in College Teaching*, 25(3–4), 57–84.

Moog, R. S., & Spencer, J. N. (2008). *Process Oriented Guided Inquiry Learning (POGIL)*, American Chemical Society Symposium Series, Vol. 994, pp. 1–13.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13–22.

National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and opportunities for 2-Year and 4-Year STEM degrees: Systemic change to support students' diverse pathways*. Washington, DC: National Academies Press.

National Research Council. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, DC: National Academies Press.

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, 2(1), 9–34.

O'Donnell, A. (2006). The role of peers and group learning. In Alexander, P., & Winne, P. (Eds.), *Handbook of educational psychology* (pp. 781–802). Mahwah, NJ: Erlbaum.

O'Donnell, A. M. (1996). Effects of explicit incentives on scripted and unscripted cooperation. *Journal of Educational Psychology*, 88, 74–86.

O'Neill, T. A., Hoffart, G. C., McLarnon, M. M., Woodley, H. J., Eggermont, M., Rosehart, W., & Brennan, R. (2017). Constructive controversy and reflexivity training promotes effective conflict profiles and team functioning in student learning teams. *Academy of Management Learning & Education*, 16(2), 257–276.

Pauli, R., Mohiyeddini, C., Bray, D., Michie, F., & Street, B. (2008). Individual differences in negative group work experiences in collaborative student learning. *Educational Psychology*, 28(1), 47–58. doi:10.1080/01443410701413746

Phelps, E., & Damon, W. (1989). Problem-solving with equals—peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology*, 81(4), 639–646. doi: 10.1037//0022-0663.81.4.639

Preszler, R. W. (2009). Replacing lecture with peer-led workshops improves student learning. *CBE—Life Sciences Education*, 8(3), 182–192. doi: 10.1187/cbe.09-01-0002

Pundak, D., & Rozner, S. (2008). Empowering engineering college staff to adopt active learning methods. *Journal of Science Education and Technology*, 17(2), 152–163. doi: 10.1007/s10956-007-9057-3

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research*, 28(6), 737–771.

Scager, K., Boonstra, J., Peeters, T., Vulperhorst, J., & Wiegant, F. (2016). Collaborative learning in higher education: Evoking positive interdependence. *CBE—Life Sciences Education*, 15(4), ar69.

Seymour, E., & Hewitt, N. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.

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

Slavin, R. E. (1991). Synthesis of research of cooperative learning. *Educational Leadership*, 48(5), 71–82.

Slavin, R. E. (2014). Cooperative learning and academic achievement: Why does groupwork work? *Anales de Psicología/Annals of Psychology*, 30(3), 785–791.

Solomon, J. (1987). Social influences on the construction of pupil's understanding of science. *Studies in Science Education*, 14, 63–82.

Solomon, Y., Croft, T., & Lawson, D. (2010). Safety in numbers: Mathematics support centres and their derivatives as social learning spaces. *Studies in Higher Education*, 35(4), 421–431.

Springer, L., Donovan, S. S., & Stanne, M. E. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51.

Stefanou, S. E., Hood, L. F., & Stefanou, C. R. (2001). Feedback and change: Assessment of individual contributions within collaborative activities in the higher education classroom. *Journal on Excellence in College Teaching*, 12(2), 77–91.

Strong, J. T., & Anderson, R. E. (1990). Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, 12, 61–67.

Svinicki, M. D., & Schallert, D. L. (2016). Learning through group work in the college classroom: Evaluating the evidence from an instructional goal perspective. In Paulsen, M. B. (Ed.), *Higher education: Handbook of theory and research* (Vol. 31, pp. 513–538). Dordrecht: Springer Netherlands.

Sweet, M., & Michaelsen, L. K. (2007). How group dynamics research can inform the theory and practice of postsecondary small group learning. *Educational Psychology Review*, 19(1), 31–47.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controveries in the use of mixed methods in the social and behvioral sciences. In *Handbook of mixed methods in social & behavioral research*, 3–50.

Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group & Organization Management*, 34(2), 170–205.

Theobald, E., Eddy, S., Grunspan, D., Wiggins, B., & Crowe, A. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS ONE*, 12(7), e0181336.

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

Van der Vegt, G. S., Emans, B. J., & Van der Liert, E. (2001). Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction. *Personnel Psychology*, 54(1), 51–69.

Wang, S.-L., & Lin, S. S. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. *Computers in Human Behavior*, 23(5), 2256–2268.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, 39(4), 943–989.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In Calfee, D. B. R. (Ed.), *Handbook of educational psychology* (pp. 841–873). Mahwah, NJ: Erlbaum.

Wenzel, T. J. (2007). Evaluation tools to guide students' peer-assessment and self-assessment in group activities for the lab and classroom. *Journal of Chemical Education*, 84(1), 182.

Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education*, *17*(1), 45–58.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. https://doi.org/10.1016/S0361-476x(02)00006-1

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, *24*(6), 420–424. doi: 10.1177/0963721415599543

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688.

Zhang, B., Johnston, L., & Kilic, G. B. (2008). Assessing the reliability of self- and peer rating in student group work. *Assessment & Evaluation in Higher Education*, *33*(3), 329–340.

Zhang, B., & Ohland, M. W. (2009). How to assign individualized scores on a group project: An empirical evaluation. *Applied Measurement in Education*, *22*(3), 290–308.