Identifying Group Work Experiences That Increase Students' Self-Efficacy for Quantitative Biology Tasks

Melissa L. Aikens* and Alexander R. Kulacki
Department of Biological Sciences, University of New Hampshire, Durham, NH 03824

ABSTRACT
Quantitative skills are a critical competency for undergraduates pursuing life science careers. To help students develop these skills, it is important to build their self-efficacy for quantitative tasks, as this ultimately affects their achievement. Collaborative learning can benefit self-efficacy, but it is unclear what experiences during collaborative learning build self-efficacy. We surveyed introductory biology students about self-efficacy-building experiences they had during collaborative group work on two quantitative biology assignments and examined how students' initial self-efficacy and gender/sex related to the experiences they reported. Using inductive coding, we analyzed 478 responses from 311 students and identified five group work experiences that increased students' self-efficacy: accomplishing the problems, getting help from peers, confirming answers, teaching others, and consulting with a teacher. Higher initial self-efficacy significantly increased the odds (odds ratio: 1.5) of reporting that accomplishing the problems benefited self-efficacy, whereas lower initial self-efficacy significantly increased the odds (odds ratio: 1.6) of reporting peer help benefited self-efficacy. Gender/sex differences in reporting peer help appeared to be related to initial self-efficacy. Our results suggest that structuring group work to facilitate collaborative discussions and help-seeking behaviors among peers may be particularly beneficial for building self-efficacy in low self-efficacy students.

INTRODUCTION
The field of biology is becoming increasingly reliant on sophisticated quantitative tools, methods, and techniques (National Research Council [NRC], 2003). These tools are required to understand and solve mounting problems in the environmental, agricultural, energy, and public health sectors (NRC, 2009). To effectively prepare life science graduates to tackle these problems, national reports have identified the importance of building students' competency in a variety of quantitative skills by incorporating these skills into the undergraduate biology curriculum (Association of American Medical College-Howard Hughes Medical Institute, 2009; NRC, 2003, 2009; American Association for the Advancement of Science, 2011). An important component to building students' competency in quantitative skills is building students' confidence in their abilities to successfully solve quantitative problems, or their self-efficacy for quantitative tasks. This is particularly relevant for life science students given the variation in perceptions of mathematical ability within this group (Sax *et al.*, 2015).

According to social cognitive theory, self-efficacy is a critical personal factor that shapes human behavior (Bandura, 1986). Self-efficacy represents an individual's beliefs and judgments about their ability to succeed at a given task (Bandura, 1997). An individual's self-efficacy for a task can influence behaviors related to the task, including the choice of whether the individual engages in the task, how much effort the individual puts into the task, and how long the individual perseveres at the task when faced with difficulties (Bandura, 1986). Engagement, effort, and persistence on challenging tasks are essential for successful performance outcomes; improving an individual's self-efficacy can thus impact performance on a given task.

Brian Sato, *Monitoring Editor*
Submitted Apr 15, 2022; Revised Jan 27, 2023; Accepted Feb 6, 2023
CBE Life Sci Educ June 1, 2023 22:ar19
DOI:10.1187/cbe.22-04-0076
*Address correspondence to: Melissa L. Aikens (melissa.aikens@unh.edu).

In an academic context, self-efficacy is posited to be a strong predictor of student motivation and performance (Bandura, 1997). Students with higher self-efficacy may have greater interest in a subject and set higher goals (e.g., Byars-Winston *et al.*, 2010). They are also more likely to use a variety of self-regulated learning techniques, such as time management and establishment of subgoals, to attain those goals (Bandura, 1997; Schunk and DiBenedetto, 2016). This, in turn, can result in increased engagement, effort, and persistence, leading to higher performance. Many empirical studies from a variety of academic contexts support the relationship between self-efficacy and student performance (reviewed by Pajares, 1996; Klassen and Usher, 2010; Schunk and DiBenedetto, 2016). For example, among undergraduate students, self-efficacy has been found to significantly predict performance in a chemistry course (Zusho *et al.*, 2003), engineering students' grade point averages (GPAs) at the end of their first year (Jones *et al.*, 2010), and underrepresented minority students' performance in a biology course (Ballen *et al.*, 2017).

Given the importance of self-efficacy to academic performance, pedagogical strategies that foster students' self-efficacy for quantitative tasks may be particularly valuable when teaching quantitative skills in biology courses. One pedagogical strategy shown to positively relate to self-efficacy is collaborative learning (Fencl and Scheel, 2005). In collaborative learning, students actively engage with one another as partners or in small groups to complete a project or assignment (Smith and MacGregor, 1992; Nokes-Malach *et al.*, 2015). For undergraduate engineering students, working in a team was reported to significantly contribute to students' self-efficacy beliefs and persistence in their academic careers (Hutchison *et al.*, 2006). However, while working in collaborative groups may be an effective strategy for increasing students' self-efficacy, the mechanism through which collaborative learning environments impact students' self-efficacy beliefs is less understood. It is not clear what experiences occur during collaborative group work that students draw upon to inform their self-efficacy beliefs or in what ways these experiences shape their self-efficacy. The current study investigates these experiences in the context of quantitative biology activities in an introductory biology course and interprets these experiences through the lens of self-efficacy theory.

Derivation of Self-Efficacy Beliefs

Bandura hypothesized that students derive their self-efficacy beliefs from four sources: mastery experiences, vicarious experiences, social persuasions (also called verbal persuasions), and physiological and affective states (Bandura, 1997; Usher and Pajares, 2008). Mastery experiences are students' perceptions of their own success after doing a task; they provide direct evidence of whether a student can complete a task and consequently are often the most influential source of self-efficacy information (Bandura, 1997; Usher and Pajares, 2008). Importantly, mastery experiences need not explicitly represent success on a task, only that students judge themselves as successful. The extent to which students judge themselves successful on a task depends on several factors, including their preconceived notions of their capabilities, the challenge of the task, the amount of effort expended on the task, and the amount of help they receive on the task (Bandura, 1997). Vicarious experiences occur through the observation of others (Bandura, 1997; Usher and Pajares, 2008). For example, students may become confident in their ability to complete a task after watching a peer similar in capability complete a task. Alternatively, students may use social comparisons—comparing their own level of success with those of their peers—as a vicarious experience. Social persuasions occur when students receive direct feedback from their peers or instructors about their performance on a task, the evaluation of which can result in self-efficacy beliefs (Bandura, 1997; Usher and Pajares, 2008). Finally, physiological and affective states relate to students' anxiety, stress, and mood induced by a task, which the students interpret as an indicator of their competence at the task (Bandura, 1997; Usher and Pajares, 2008).

Each of the sources of self-efficacy are weighed and integrated to form a student's self-efficacy beliefs (Bandura, 1997). In some cases, one source may primarily contribute to the development of self-efficacy. This source may be deemed a more reliable indicator of capability than other sources or provide new information about a student's capabilities, and thus have a greater influence on the development of self-efficacy beliefs (Bandura, 1997). Alternatively, multiple sources may contribute to the development of self-efficacy beliefs (Chen and Usher, 2013). Multiple sources of self-efficacy may even be elicited from a single experience due to the interwoven nature of the sources of self-efficacy (Usher and Pajares, 2008; Morris and Usher, 2011; Usher *et al.*, 2019). As an example, a student may obtain good grades in math and perceive that as a mastery experience while simultaneously recognizing their grades are higher than most other students' grades, which is a vicarious experience (Usher *et al.*, 2019). Quantitative studies have found high correlations between the sources of self-efficacy, supporting the view that the sources may often be associated with one another (Lent *et al.*, 1991; Byars-Winston *et al.*, 2017; Sheu *et al.*, 2018).

The ways in which students interpret their experiences and perceive and weigh the sources of self-efficacy can vary greatly among individuals. These differences may be due to cultural differences among students (Ahn *et al.*, 2016), but the way in which a student perceives and weighs a source of self-efficacy also depends on the contextual features of the experience (Bandura, 1997). For example, students may vary in their perceptions of the difficulty of a task, and the perceived difficulty of the task can influence the extent to which students weigh a mastery experience in their self-efficacy formation (Bandura, 1997). Additionally, the perceived similarity of a role model can influence the extent to which a vicarious experience informs the development of self-efficacy, and the perceived knowledgeability and sincerity of an individual providing encouragement can influence the extent to which a social persuasion informs the development of self-efficacy (Bandura, 1997). Different contexts, such as academic subjects, may also invoke different physiological and affective responses that can influence a student's self-efficacy. For example, middle and high school students, on average, reported higher physiological and affective states (e.g., stress) for math than for science, resulting in physiological and affective states significantly predicting math self-efficacy, but not science self-efficacy (Usher *et al.*, 2019). As Bandura (1997, p. 115) noted, the formation of self-efficacy beliefs is "a complex process of self-persuasion."

Building Students' Self-Efficacy through Collaborative Group Work

Although students may build their self-efficacy from any number of classroom experiences (e.g., mastery of an individual homework assignment), collaborative group work may be uniquely positioned to build students' self-efficacy. The practice-oriented, interactive nature of group work may lend itself to generating a breadth of experiences that could build self-efficacy in multiple ways. The simple act of completing a problem on a group work assignment can engender mastery judgments by providing students with an opportunity to succeed at the task. However, the social interactions that occur during group work have the potential to influence students' self-efficacy through vicarious experiences and social persuasions. When working collaboratively, students may observe the success of their peers on a problem (Nokes-Malach *et al.*, 2015), which may stimulate confidence in their own abilities to successfully answer the problem (Sawtelle *et al.*, 2012). Discussions among group members may also include support toward solving a problem and encouraging feedback about solutions, promoting positive social persuasions (Hutchison *et al.*, 2006; Purzer, 2011). Additionally, groups provide opportunities for students to seek help from one another, which students have reported to be an important factor influencing their self-efficacy (Hutchison *et al.*, 2006; Butz and Usher, 2015). Importantly, one group work experience may relate to multiple sources of self-efficacy. For example, positive feedback from other students on the way in which a student solved a problem may result in an increase in self-efficacy through both social persuasions from peers and a mastery experience. Of course, it is also possible that group work could decrease a student's self-efficacy. If no students in the group can solve the problem, then no mastery experience would be achieved, or a negative vicarious experience could occur if any of the students judged themselves unfavorably compared with other group members. Presently, it is unclear which experiences during group work are particularly salient for the development of self-efficacy beliefs. Moreover, students may weigh experiences differently in the formation of their self-efficacy beliefs based on personal factors (Usher and Pajares, 2008).

One important personal characteristic that might influence which experiences students primarily draw from to inform their self-efficacy beliefs is gender. Although some studies have found that both men and women predominantly rely on mastery experiences (Matsui *et al.*, 1990; Lent *et al.*, 1991; Britner and Pajares, 2006), other studies have found gender differences in how the sources influence students' self-efficacy (e.g., Byars-Winston *et al.*, 2017). Recent qualitative studies have reported that women more frequently identify vicarious experiences (Butz and Usher, 2015; Webb-Williams, 2018), social persuasions (Butz and Usher, 2015; Usher *et al.*, 2019), or help availability (Hutchison *et al.*, 2006; Usher *et al.*, 2019) as important to the development of their self-efficacy than men. Collectively, these studies suggest women may be more likely to use social-based sources of self-efficacy than men. In the context of group work, this would imply that experiences involving peer interactions would be more influential in the development of women's self-efficacy than men's self-efficacy. Ultimately, if men and women rely on different sources to inform their self-efficacy beliefs, then different group work experiences will be salient for men and women in the development of their self-efficacy beliefs for quantitative biology tasks.

Students' initial self-efficacy can also influence the experiences they evaluate when building their self-efficacy. The self-efficacy for a task that students bring into the group setting can affect how they behave and, thus, the experiences they have in the group. Students with high self-efficacy may be more likely to answer questions and share ideas than students with low self-efficacy (Purzer, 2011). Alternatively, if a student's self-efficacy is accurately calibrated to their ability, then students with high self-efficacy may instead prefer to work alone on the problem (e.g., French *et al.*, 2011) and participate sparingly in discussion with group members. There are few data on how students' initial self-efficacy affects the experiences or sources they draw from to evaluate their self-efficacy. Recent work found that students with high self-efficacy were more likely to report that mastery experiences, vicarious experiences through social comparisons, or social persuasions increased their self-efficacy, whereas students with low self-efficacy were more likely to report nothing increased their self-efficacy (Butz and Usher, 2015). However, this study did not assess initial self-efficacy before an activity aimed at building self-efficacy.

A student's initial self-efficacy for a quantitative biology task is also likely related to gender. Negative stereotypes about women in mathematics can undermine women's confidence in their mathematical abilities (Schmader *et al.*, 2004; Franceschini *et al.*, 2014). Indeed, several studies have reported that women have lower self-efficacy for mathematics than men (reviewed by Pajares, 2005). Thus, gender differences in the sources of self-efficacy used to form self-efficacy beliefs may in part be related to different behaviors in which men and women engage due to the different self-efficacy beliefs they bring into a task. Incorporating both gender and initial self-efficacy into analyses can provide insight into what extent gender differences in the sources of self-efficacy accessed are due to differing initial self-efficacy levels.

In summary, peer interactions through collaborative group work provide many opportunities for self-efficacy beliefs to develop. Therefore, group work may be an important pedagogical tool for improving students' confidence in their ability to solve quantitative problems. However, we lack data on the specific group work experiences that different students find particularly salient in developing their self-efficacy beliefs. In this study, we identify group work experiences that students report as increasing their self-efficacy for quantitative biology tasks and relate these experiences to student gender and initial self-efficacy. We focus on identifying group work experiences that are sources of self-efficacy for students, rather than identifying Bandura's (1997) theoretical sources of self-efficacy, because these are of practical importance to instructors, as they represent behaviors that can be encouraged. Moreover, one group work experience can relate to multiple of Bandura's sources of self-efficacy, making it difficult to ascribe any one of Bandura's sources to a given experience. Instead, we elaborate on how the experiences reported by students may relate to Bandura's sources of self-efficacy in the *Discussion*.

Research Questions

We sought to better understand how collaborative learning through group work can increase a student's self-efficacy for quantitative biology tasks. Specifically, we addressed the following questions: 1) What self-reported experiences during group work do students believe increase their self-efficacy for quantitative biology tasks? 2) How does students' gender and their initial self-efficacy for quantitative biology problems relate to the experiences they report as increasing their self-efficacy for quantitative biology tasks? Identification of group work experiences that increase students' self-efficacy—and how these experiences vary among students—can be used to inform the design of group work in a way that maximizes these experiences to promote student motivation and learning.

METHODS

Setting and Participants

We surveyed undergraduate students in two sections of an introductory biology course, each taught by a different instructor, at a large public research university in the northeastern United States that is a primarily White institution. This course is one of two required introductory biology courses for many life science majors at this institution, although non–life science majors may take this course to fulfill a general education requirement. The curriculum of the course covers a variety of topics in evolution, biodiversity, and ecology, such as phylogenetics, the mechanisms and principles of evolution, biological speciation, the major plant and animal groups, population ecology, and community ecology.

In-class group work was a component of both sections of the course (hereafter referred to as course sections A and B). At the beginning of the semester, students were assigned to groups of three to five students with an average group size of four students in both courses. Groups were primarily assigned according to students' seating preferences, such as preferring to be near the front of the room or off to one side. However, in group formation, the instructors tried to also ensure that groups would not have a single woman or a single man. These groups sat together during class sessions and remained together throughout the semester. Groups would work together on average once a week to complete a collaborative in-class assignment based on lecture content; one assignment would be turned in per group. In course section A, one randomly chosen student in each group was assigned to be a scribe for the assignment, but no other roles were assigned. Students were not assigned a specific group role in course section B. A graduate teaching assistant graded (for effort in course section A; for accuracy in course section B) and provided feedback on the group work assignments. Concepts from the group work assignments were incorporated into summative assessments for both sections of the course.

Participants in each course section were asked to complete two separate surveys about two different group work assignments for a small amount of course credit, though participation in the research study was optional. Course section A had 200 students, and course section B had 161 students. Of these students, 173 (87%) and 61 (38%), from each section respectively, attended class on the day of the first group work assignment, completed the first survey, and consented to participate in the research, and 121 (61%) and 123 (76%) attended class on the day of the second group work assignment, completed the second survey, and consented to participate in the research. This resulted in a total of 478 responses across the two surveys from 311 unique students. Participants' demographic information is summarized in Table 1. This study was approved by the Institutional Review Board at the authors' institution (IRB no. 7005).

Data Collection

Our study centers around two in-class group work assignments: one in which students evaluated Hardy-Weinberg equilibrium (HWE), and another in which students modeled exponential and logistic population growth (PG). The HWE assignment differed between the sections. In course section A, students completed a modified case study in which they had to apply HWE to a scenario about the conservation of timber rattlesnakes (modified from Drott and Sarvary, 2016). In course section B, students completed an assignment in which they used HWE to evaluate whether aposematic coloration traits in burying beetles were evolving. For PG, the assignment for both sections was the same; it asked students to calculate the population growth rate of invasive Burmese pythons in the Everglades under exponential growth conditions and to predict the population size of invasive brown tree snakes in Guam under logistic growth conditions. In both course sections, the HWE assignment occurred early in the semester (week 3 in course section A, week 4 in course section B), while the PG assignment occurred in the latter half of the semester (week 12 in both sections).

On the day of each of the group work assignments described, we provided students with a paper-based survey at the beginning of class (the pre survey), before starting the group work. Each pre survey consisted of a sample quantitative problem. The HWE pre survey provided students with a table showing the number of individuals of each genotype in a population, and asked students to report their confidence in their ability to do two things with this problem using a five-point scale (ranging from "1–Not at all confident" to "5–Completely confident"): 1) calculate the predicted number of individuals in the population under HWE conditions and 2) justify whether the population is evolving or not (Supplemental Material). The PG pre survey provided students with demographic information about a cod population and asked students to report their confidence (ranging from "1–Not at all confident" to "5–Completely confident") in their ability to predict the population size of cod at a future time point using the logistic growth equation (Supplemental Material). Students were not expected to actually answer the questions on the pre surveys and, in fact, were not given enough time to do so. After we collected the pre survey, students worked together in their groups to complete the collaborative assignment. The instructors of each section as well as a graduate teaching assistant (in both course sections) and several undergraduate teaching assistants (in course section A only) circulated throughout the room to answer any questions that arose during group work.

Following the end of the class session, we administered a post survey to students using the online service Qualtrics, delivered via the course's online learning management system. To address our research questions, we analyzed a subset of questions from this post survey that broadly queried students' group work experiences. These questions included students' confidence in their ability to answer the same questions from the pre survey using the same five-point scale and the open-response question: "Describe any experiences and/or interactions during group work today that increased your confidence in your ability

TABLE 1. Participant demographics (n = 311)[a]

	Percentage of participants in course section A	Percentage of participants in course section B	Percentage of all participants
Gender/sex[b]			
Male	33%	29%	32%
Female	66%	69%	67%
Other	1%	0%	<1%
Year in school			
First year	72%	71%	72%
Second year	17%	12%	15%
Third year	7%	13%	10%
Fourth year	3%	2%	2%
Fifth year	1%	1%	1%
Highest high school math			
Algebra or geometry	13%	7%	11%
Trigonometry	7%	9%	7%
Pre-calculus	40%	53%	46%
Calculus	38%	30%	35%

[a]Percentages may not total to 100% due to students who chose "Prefer not to respond."
[b]Although we queried students about their gender identity, the response options reflected sex rather than gender. Therefore, we refer to this demographic variable as "gender/sex."

to [solve the sample problem]" (Supplemental Material). Although we also asked students to report experiences that decreased their confidence in their ability to solve quantitative biology problems, an error in survey deployment prevented us from being able to report student quotes related to negative experiences under the IRB. At the end of the post surveys, we also asked students, "With which gender do you identify?" However, we provided the options "male," "female," "other" (with a write-in text box), and "prefer not to respond," thereby conflating gender with sex. Because of this conflation, we hereafter use the term "gender/sex" when referencing these results. Additionally, because students identified with "male" or "female," we report our results in terms of male students and female students, rather than men and women. We recognize that this survey item is not inclusive with respect to gender, because it permits the othering of individuals with queer-spectrum identities (Cameron and Stinson, 2019; Casper *et al.*, 2022). We discuss this further in the *Limitations* section of the *Discussion*.

In the first surveys about HWE, we intentionally asked students about their confidence to answer two types of questions: one that only involved calculations ("Calculate predicted number of individuals…") and one that involved evaluation ("Justify whether the population is evolving…"; Supplemental Material). However, during preliminary analyses, it became apparent that students were responding similarly in their confidence ratings to both the calculation and the evaluation questions. In other words, students did not appear to be differentiating between the two types of questions. Therefore, we only analyzed the calculation problem on the HWE pre and post surveys, and we only included a calculation problem on the PG pre and post surveys.

We created both the rating scale items and the open-response item, using previous studies as a guide. Although Likert scales validated as measures of self-efficacy exist, we created our own items to measure self-efficacy, because measures of self-efficacy need to be "tailored to the particular domain of functioning that is the object of interest" (Bandura, 2006). Our interest was in understanding students' confidence in their ability to solve the specific quantitative problems being practiced in group work. However, existing self-efficacy instruments address self-efficacy more broadly, for example, by measuring students' confidence in their ability to use biology methods and apply biology concepts and skills (Biology Self-Efficacy Scale; Baldwin *et al.*, 1999), to do well in a course (Motivated Strategies for Learning Questionnaire; Pintrich and De Groot, 1990), to perform accomplishments required for success in science and engineering majors (strength of self-efficacy for academic milestones; Lent *et al.*, 1986), and to do well in science generally (Science Motivation Questionnaire II; Glynn *et al.*, 2011). Therefore, for our purposes, we created a task-specific self-efficacy item similar in structure to task-specific items developed to measure mathematics self-efficacy, where students are given a specific math problem and asked how confident they are that they can solve it (Bandura and Schunk, 1981; Pajares and Graham, 1999). However, unlike the math task–specific measures, in which students rated their self-efficacy to solve a variety of math problems, each of our measures included only one problem. We opted to include only one item because we were measuring a concrete construct that was narrow in focus (confidence in ability to do a specific problem; Fuchs and Diamantopoulos, 2009), rather than something more complex, such as self-efficacy for quantitative biology skills or mathematical skills. We loosely modeled our open-response item on an item used in a previous study by Butz and Usher (2015). However, we tailored the wording of the item to specifically query any experiences and/or interactions that happened in group work to increase self-efficacy. We address the limitations of our items in the Discussion section.

Data Analyses

This study uses a concurrent transformation mixed-methods design (Warfa, 2016). Qualitative and quantitative data were collected during a single phase, as described earlier. Data analysis began by conducting a qualitative analysis of students' short responses. The resulting codes were then quantitated into dichotomous variables to be used in subsequent statistical analyses. We also statistically analyzed change in students' pre and post self-efficacy scores to examine overall changes in self-efficacy.

Analysis of Self-Efficacy Scores. We conducted separate statistical tests on the HWE and the PG data because students may be represented in both data sets, violating assumptions of independence. The Shapiro-Wilk test for normality indicated self-efficacy scores were not normally distributed ($p < 0.001$ for both HWE and PG data); this was visually confirmed by examining histograms of the data. Therefore, we used paired Wilcoxon ranked-sign tests to determine whether self-efficacy significantly changed from the pre to the post survey. Only students who had both pre and post self-efficacy scores were included in these analyses ($n = 230$ for the HWE data; $n = 234$ for the PG data).

Qualitative Coding of Short Responses. We used an inductive approach to identify group work experiences that students reported increased their self-efficacy. We inductively coded group work experiences, because we were first and foremost interested in identifying tangible practices that occurred during group work that have practical relevance to instructors seeking to maximize positive student outcomes from group work. In coding students' experiences, we used process coding, which involves using gerunds to describe events, occurrences, or ongoing action in a situation of interest (Saldaña, 2016). Process coding was especially useful for our study, because we were interested in capturing specific moments and actions during students' overall group work experience that may have impacted their self-efficacy (Saldaña, 2016).

We started our analysis by reviewing students' responses to get a general sense of the types of experiences students described. Three members of the research team (A.R.K. and two undergraduate research assistants) parsed all student responses to develop a preliminary codebook of group work experiences that increased students' self-efficacy. Throughout this preliminary coding, the team used analytic memos to organize thoughts, identify patterns and notable responses, and develop a deeper, more holistic understanding of the responses to ensure consistency in the codebook (Saldaña, 2016). Once the preliminary codebook of group work experiences was established, two members of the research team (M.L.A. and A.R.K.) conducted an iterative coding process.

We first independently coded two "training rounds" of 40 responses each, drawn from both HWE and PG surveys and from both sections of the course. Because one coder (A.R.K.) created the training round responses from the post-survey data, that coder was not blind to the section of the course the responses came from. However, the second coder (M.L.A.) was blind to the section of the course each response came from. Following each training round, we discussed the codes we assigned to each response and resolved any disputes or disagreements to come to consensus on all codes. Based on our discussions, we combined codes or revised definitions of codes. For every revision to the codebook, we would re-examine previously coded responses to ensure the revised code or definition was consistent with our data. Following these two training rounds and the solidification of the codebook, the research team independently coded an additional third and fourth round of 40 responses each, following a similar pattern to the training rounds, while also assessing interrater reliability (IRR) for each of our developed codes. Per the recommendations of Xu and Lorber (2014), we established Holley and Guilford's G-index (Holley and Guilford, 1964) as our metric for IRR, because of its general robustness with skewed responses, which we expected our data to exhibit given that some codes were not common within the 40 responses of a round. We established thresholds for achieving IRR based on the recommendations of Hruschka and colleagues (2004): index of agreement (G) > 0.80 for most (>90%) of the codes. We calculated G-indices for the third and fourth rounds of coding and achieved a $G > 0.80$ across 95% of codes, with a minimum G-index of 0.75, which occurred on only one code in the third round of coding. Following this, one member of the research team (A.R.K.) independently coded the remaining experiences that increased self-efficacy. After completion, a random selection of 30% of these remaining coded responses was shared with a member of the research team (M.L.A.) to confirm IRR. We also met our threshold for IRR for this last set of codes, with a minimum G-index of 0.93.

After coding all responses, we examined the codes to determine whether some could be combined into broader categories. Of the nine codes we found, we decided that three codes—Discussing/Working Together (talking and working through the problem as a group to solve it), Being Taught/Guided (being explicitly taught or guided through the problem by their peers), and Asking Questions (actively asking questions of other group members)—were very similar, in that they all represented forms of Getting Help from Peers, and we therefore categorized these three codes under this larger category. We also considered how the group work experiences aligned with the sources of self-efficacy. Although some experiences closely aligned with Bandura's (1997) sources, other experiences were more difficult to clearly categorize as one specific source of self-efficacy. In fact, we considered how some experiences may relate to multiple sources of self-efficacy. Therefore, we opted not to categorize each experience as a source of self-efficacy, but rather reserve our interpretation of how the experiences may relate to one or more sources of self-efficacy for the *Discussion*.

We coded a total of 478 responses. Of the responses for which we were able to match pre–self-efficacy scores (464 responses), 34 responses (7%) indicated a decrease in self-efficacy scores from the pre survey to the post survey. However, the majority of these responses reported a group work experience that increased their self-efficacy. We included these responses in all of our analyses, because students' self-efficacy can both increase and decrease over the course of group work, meaning that self-efficacy-building experiences can occur even if overall self-efficacy from the pre survey to the post survey decreases. For example, students who are overconfident in their abilities may have a significant decline in their self-efficacy as they begin a task but may have a self-efficacy-building experience before the end of the task. This experience may increase their self-efficacy, though perhaps not enough to reach the self-efficacy levels initially reported. For this reason, we did not want to exclude students whose self-efficacy decreased from the pre survey to the post survey.

Statistical Analyses of Group Work Experiences. We first tested for differences in initial self-efficacy scores between male students and female students. Separate analyses were conducted for HWE self-efficacy scores and PG self-efficacy scores to avoid violating the assumption of independent observations. Shapiro-Wilk tests indicated the data violated assumptions of normality ($p < 0.001$ for both genders/sexes in both the HWE and PG data sets); this was visually confirmed by examining histograms of initial self-efficacy scores by gender/sex in each data set. Therefore, we used the nonparametric Mann-Whitney *U*-test to detect significant differences in initial self-efficacy by gender/sex.

To address our second research question, we statistically modeled how initial self-efficacy and gender/sex predict the probability of reporting a specific experience increases self-efficacy. Because few students reported identifying with a gender/sex other than male or female (<1%), we were only able to include students who identified as a male student or a female student in our statistical analyses that included gender/sex as a variable due to sample size constraints. Additionally, only student responses that included a pre–self-efficacy score and a group work experience that could be coded (i.e., not coded as Non-answer) were included in the statistical analyses ($n = 429$). We were limited to statistically modeling only three experiences, Accomplishing It, Getting Help from Peers, and Confirming Their Answers, as other experiences were reported too infrequently (fewer than 35 responses reported the experience out of 429 responses) to be able to draw robust conclusions about how initial self-efficacy and gender/sex relate to these experiences. When there are few observations for a combination of predictor variable and outcome levels (i.e., sparse data), regression coefficients can be biased (Greenland *et al.*, 2016). Moreover, when we tried to run the models that contained infrequent responses, they failed to converge and therefore did not yield valid results.

We used generalized linear mixed models to examine the three experiences (Accomplishing It, Getting Help from Peers, and Confirming Their Answers). Generalized linear mixed models allow us to model probability outcomes from predictor variables while simultaneously controlling for other, unintended variables that may affect the probability outcomes (e.g., question context, course section), correlations between measurements from the same students (repeated measures), and correlations due to the nested structure of students within groups. Our response variable in each model was expressed as a binary outcome of whether students reported the experience in their survey responses (0 if they did not report the experience, 1 if they did report the experience). Therefore, we used a logit link function in our models, which represents a mixed-effects logistic regression model.

For each group work experience, we ran three regression models: one with gender/sex as a predictor variable, one with gender/sex and initial self-efficacy as predictor variables, and one with the interaction between gender/sex and initial self-efficacy as a predictor variable. This allowed us to investigate the extent to which any observed gender/sex differences in reporting a group work experience relate to initial self-efficacy. In all models, question type (HWE or PG) was included as a control variable for two reasons. First, self-efficacy is task specific (Pajares, 1996), and thus students may rely on different experiences to increase their self-efficacy in different contexts. Second, the two questions occurred at different points in the semester, and as students worked with their group over the semester and gained greater familiarity with one another, they may have changed how they weighted different experiences in building their self-efficacy. Course section (A or B) was also included as a control variable in all models to account for variation in the group work experiences students reported due to different classroom climates and structures. We also included student (represented by a unique ID for each student and used to account for repeated measures) as a random effect in each model. We initially ran every model with group (represented by a unique ID for each group and used to account for clustering of students within semester-long groups) as a random effect, too. However, whenever group was included as a random effect, it resulted in fitted models that were singular. Inspection of the variance estimates for group revealed that the variance for the group random effect was zero or nearly zero in all cases. Therefore, we removed group as a random effect from all models, as it did not appear to be accounting for any of the variation in the models. The final R models, run for each of the three group work experiences we statistically analyzed, were:

1. report of group work experience (0/1) ~ gender/sex + question type + course section + (1|student)
2. report of group work experience (0/1) ~ initial self-efficacy + gender/sex + question type + course section + (1|student)
3. report of group work experience (0/1) ~ initial self-efficacy*gender/sex + question type + course section + (1|student)

All statistical analyses were conducted using R v. 4.1.2 (R Core Team, 2021). We used the rstatix package (Kassambara, 2022) to calculate effect sizes for the change in self-efficacy scores. We used the lme4 package (Bates *et al.*, 2015) to run the generalized linear mixed models. We calculated a marginal R^2 and a conditional R^2 for each model using the delta method for bias correction in binomial generalized linear mixed-effects models (Nakagawa *et al.*, 2017) using the piecewiseSEM package (Lefcheck, 2016). Marginal R^2 represents the total variance explained by only the fixed effects (initial self-efficacy, gender/sex, question type, and course section), and conditional R^2 represents the total variance explained by both fixed and random effects.

RESULTS

Students' self-reported self-efficacy for both the HWE task and the PG task significantly increased from the pre to the post survey ($p < 0.001$ for both data sets; Figure 1 and Supplemental Table S1). Mean self-efficacy scores increased from 3.35 (SE = 0.20) to 3.92 (SE = 0.17) for the HWE task and from 2.85 (SE = 0.20) to 3.34 (SE = 0.18) for the PG task. The effect size was large in both cases ($r = 0.58$ and $r = 0.51$, respectively). Therefore, on average, students were more confident in their ability to solve the HWE and PG problems after completing a group work assignment. However, given the limitations of our measure of self-efficacy, we focus on the qualitative data to understand what group work experiences increased self-efficacy.

We coded 234 student responses for the HWE open-response question and 244 student responses for the population growth open-response question. We identified seven process codes for

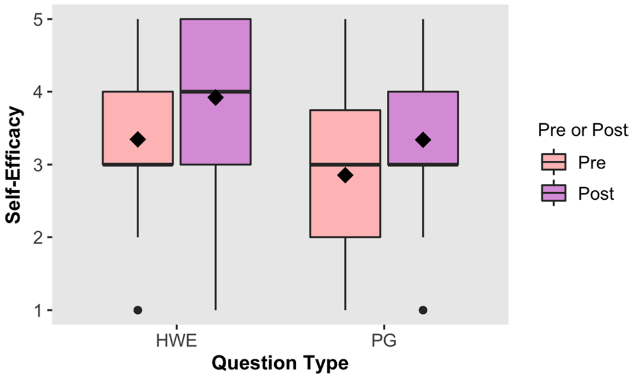

FIGURE 1. Distribution of pre and post self-efficacy scores for each question type. Mean scores are represented by the black diamonds. Only students who completed both the pre and post surveys are included (*n* = 230 for HWE; *n* = 234 for PG).

experiences that increased self-efficacy (Table 2): Accomplishing It, Discussing/Working Together, Being Taught/Guided, Asking Questions, Confirming Their Answers, Teaching/Guiding Others, and Consulting with a Teacher. We additionally identified two "neutral" codes: No Impact (6% of student responses; 5% of total codes assigned), for when students expressed that no experiences occurred during group work which impacted their self-efficacy; and Non-answer (6% of student responses; 6% of total codes assigned), for when students provided a blank, incomplete, unintelligible, or irrelevant response, bringing the total number of identified codes to nine. Student responses in which multiple experiences were discussed (54 responses) were coded for each experience. This resulted in a total of 539 codes assigned across the 478 student responses. Representative quotes included in the *Results* and *Discussion* have been lightly edited for spelling and grammar.

One of the most commonly reported group work experiences that increased students' self-efficacy for solving the quantitative problems was Accomplishing It (Figure 2), which represents a feeling of competence gained directly from completing the problems in the group work assignment. A total of 29% of student responses included a reference to Accomplishing It, and this code represented 26% of total codes assigned. Accomplishing It includes students who mention practicing the quantitative biology problems, working through the assignment, or doing the assignment. For example, one student responded, "The group work was helpful because it provided me with additional practice." Other students made more explicit references to the importance of "doing" a problem: "Actually doing several calculations increased my confidence." Responses coded as Accomplishing It did not include references to relying on group members to be able to solve the problems. Rather, this code embodied how the simple act of doing a problem increased self-efficacy.

Experiences in which students relied on peers to help them solve the problems were also commonly reported. We identified three experiences, Discussing/Working Together, Being Taught/Guided, and Asking Questions, that we collectively categorized as Getting Help from Peers (Table 2). This category, found in 45% of student responses, represents how help or guidance from a student's peers during the problem-solving process contributes to a student's self-efficacy beliefs. Discussing/Working Together, which represents the benefit of simply being able to talk to group members, discuss ideas, and work through the problems together, was the most commonly described experience of help availability from peers (included in 28% of student responses; 25% of total codes assigned; Figure 2). Students

TABLE 2. Group work experiences, individual codes, individual code definitions, and representative quotes for each code

Group work experience/code		Definition	Example quote
Accomplishing It		The student feels confident after practicing problems as part of the group work session.	"Practicing using given data to calculate growth rates and applying carrying capacity to the logistic growth model."
Getting Help from Peers	Discussing/Working Together	The student feels confident after talking through the problem and working together to solve it.	"Being able to work with someone helped because we could share ideas on how to solve the problems."
	Being Taught/Guided	The student feels confident after being explicitly taught how to do the problem or by having someone walk them through the steps of the problem.	"My partner that sits right next to me showed me in depth step by step that helped me with understanding the HWE."
	Asking Questions	The student feels confident after actively asking questions of other group members when they did not understand something.	"Being able to ask another member of the group how a certain part was done so that I didn't get stuck."
Confirming Their Answers		The student checks their own answers with other members of the group and feels more confident in their own answer.	"Calculating the same answers as my other group mates."
Teaching/Guiding Others		The student feels confident because they are able to explain the answers to others in the group.	"Teaching other group members or other groups what to do."
Consulting with a Teacher		The student feels confident after receiving help from an instructor or teaching assistant.	"After talking to our TA she was able to help explain how the whole equilibrium works so I feel more confident doing the equilibriums."

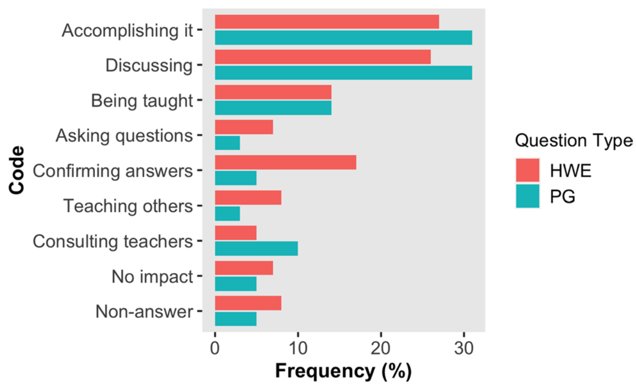

FIGURE 2. Frequency of responses that report each group work experience for each question type. Frequencies for each question type (*n* = 234 for HWE; *n* =244 for PG) add up to > 100% because some students reported multiple group work experiences in their responses. The codes Discussing, Being Taught, and Asking Questions were consolidated into the category of Getting Help from Peers.

whose responses were coded as this experience indicated that simply being able to work in a group and talk to their peers was beneficial, as it allowed them to share ideas and thoughts. For example, one student explained how the discussions helped her see the problem from different perspectives: "The discussion amongst my group members helped me to see the problems from different angles and how best to approach and complete each." Other students reported that receiving help—either through group members who offered to guide them through challenging parts of the assignment or through an ability to ask targeted questions—was important for increasing their self-efficacy for quantitative biology problems. We used two codes to represent these forms of help: Being Taught/Guided, which represents experiences when students reported receiving clarification or help from their group members on the assignment, and Asking Questions, which represents responses in which students specifically reported being able to actively ask questions of their peers when help was needed. Being Taught/Guided was coded in 14% of student responses (12% of total codes assigned) and Asking Questions was coded in 5% of student responses (4% of total codes assigned). Being Taught/Guided did not necessarily indicate a student explicitly asked group members for help, but rather was related to more general statements of receiving help, such as: "When I was confused or stuck on a certain step, my group members explained to me what I was doing incorrectly and helped me understand the math." In contrast, responses coded as Asking Questions specifically referenced the importance of being able to actively ask questions when help was needed: "Being able to ask questions to my peers really helped me understand the reasons why we were doing some of the steps."

Students also relied on their peers to check their answers, which they reported increased their self-efficacy. We used the code Confirming Their Answers (included in 11% of responses; 10% of total codes assigned; Figure 2) when students reported that checking, confirming, or comparing their answers to their group members' answers, or referencing that they achieved the same answer as their group members, increased their confidence in their ability to do the quantitative problem. This student explained how checking answers confirmed she was doing the problems correctly: "When other people in my group and I got the same answers it helped reinforce that what I was doing was right." We differentiate this experience from Getting Help from Peers, because it did not involve actively relying on peers to get through the problem-solving process, but rather looking to peers after completing a problem to ensure an answer was correct.

While many students reported receiving help from peers increased their self-efficacy, a small number of students reported that giving help increased their own self-efficacy. Only 6% of student responses (5% of total codes assigned) included the code Teaching/Guiding Others, in which students reported a considerable increase in confidence in their own ability to solve the problems because they were able to teach or guide their peers about the topic of the assignment (Figure 2). For example, one student simply stated, "I was able to answer questions and assist other members of my group by explaining the answers to them." Another student explicitly stated that teaching others helped in understanding the concepts better: "Helping explain it to people who were confused reinforced it for me."

Finally, we identified a distinct experience of receiving help from teachers (Table 2): Consulting with a Teacher. This code was present in 8% of responses (7% of total codes assigned). It represents experiences when students received help from an instructor, graduate teaching assistant, or undergraduate teaching assistant: "The one-on-one interaction with the professor helped clarify my understanding."

Relationships between Initial Self-Efficacy, Gender/Sex, and Group Work Experiences That Increase Self-Efficacy

Female students reported lower self-efficacy, on average, than male students (Supplemental Table S2). Mean initial self-efficacy scores for the HWE and PG problems were significantly lower for female students than male students ($p = 0.05$ and $p < 0.01$, respectively), although effect sizes were small in both cases ($r = 0.14$ and $r = 0.19$, respectively). Mean initial self-efficacy scores for the HWE task were 3.27 (SE = 0.09) for female students and 3.59 (SE = 0.12) for male students. Mean initial self-efficacy scores for the PG task were 2.75 (SE = 0.08) for female students and 3.21 (SE = 0.15) for male students. Therefore, comparing a model with gender/sex as the lone predictor to a model with both gender/sex and initial self-efficacy as predictors can provide insight into the extent to which differences between genders/sexes in reporting a group work experience may be caused by differences in initial self-efficacy.

We intended to examine how initial self-efficacy and gender/sex affect the probability of reporting each experience, but some experiences were reported too infrequently to statistically analyze (Table 3). Specifically, Teaching/Guiding Others, Consulting with a Teacher, and No Impact were reported in fewer than 35 out of 429 responses. Examining the descriptive statistics in Table 3, the raw means for initial self-efficacy differ between students who reported these experiences and students who did not. Students who reported Teaching /Guiding Others and No Impact had higher initial self-efficacy, on average, than students who did not report these experiences. Students who reported Consulting with a Teacher had lower initial self-efficacy, on average, than

TABLE 3. Number of responses (*n*), mean initial self-efficacy, and the percent of male students (% M) and female students (% F) that correspond to responses in which an experience was absent (0) or present (1)

		n	Mean initial self-efficacy (±SD)	% M	% F
Accomplishing It	0	292	3.02 (1.07)	29	71
	1	137	3.35 (1.11)	35	65
Getting Help from Peers	0	225	3.32 (1.19)	36	64
	1	204	2.91 (0.92)	25	75
Confirming Their Answers	0	377	3.06 (1.09)	31	69
	1	52	3.62 (0.95)	37	63
Teaching/Guiding Others	0	402	3.09 (1.09)	30	70
	1	27	3.70 (0.99)	48	52
Consulting with a Teacher	0	395	3.21 (1.06)	33	67
	1	34	2.18 (0.97)	9	91
No Impact	0	403	3.10 (1.07)	30	70
	1	26	3.58 (1.39)	54	46

students who did not report this experience. Additionally, male students are more likely, relative to their proportion of the sample, to report Teaching/Guiding Others and No Impact, whereas female students are more likely, relative to their proportion of the sample, to report Consulting with a Teacher. However, generalized linear mixed-effects models with these experiences as outcomes failed to converge, likely due to the small number of events per variable. Therefore, in addressing our second research question, we focused our analyses on the three experiences that were reported more frequently: Accomplishing It, Getting Help from Peers, and Confirming Their Answers.

Initial self-efficacy, but not gender/sex, predicted the probability a student reported that Accomplishing It increased self-efficacy. For a one-unit increase in self-efficacy levels, holding all other variables constant, the odds of a student reporting Accomplishing It increase by a factor of 1.5 (Table 4). Therefore, higher initial self-efficacy is related to a greater probability of reporting that simply completing the problems during group work increased self-efficacy for the quantitative biology problems (Figure 3). Gender/sex was not significant as a main effect or as part of an interaction term in any of the models for Accomplishing It (Supplemental Table S3).

Both gender/sex and initial self-efficacy are related to the probability of students reporting Getting Help from Peers increased their self-efficacy (Table 4). When gender/sex is the lone predictor in the model, it significantly predicts the probability of a student reporting Getting Help from Peers. The odds of a female student reporting Getting Help from Peers is 1.9 times greater than the odds of a male student reporting Getting Help from Peers. However, when initial self-efficacy is entered into the model, initial self-efficacy significantly predicts Getting Help from Peers, but gender/sex is no longer significant. For a one-unit decrease in self-efficacy levels, holding all other variables constant, the odds of a student reporting a peer help availability experience increase by a factor of 1.6. Therefore, lower initial self-efficacy is related to a greater probability of reporting that a peer help availability experience during group work increased self-efficacy for the quantitative biology problems (Figure 4). The interaction term between gender/sex and initial self-efficacy was not significant (Supplemental Table S4).

Neither initial self-efficacy nor gender/sex were related to the probability of reporting Confirming Their Answers (Supplemental Table S5). Given the relatively large standard errors for the regression coefficients in these models, it is likely that the 52 reported experiences of Confirming Their Answers (Table 3) were too few to adequately and reliably model the fixed effects and random effects included in the models. Moreover, we were unable to test the interaction term, because this model would not converge, again likely due to too few experiences reported. Therefore, we recommend that this experience, along with Teaching/Guiding Others, Consulting with a Teacher, and No Impact, be examined in larger studies where more of these events are likely to be reported.

DISCUSSION

Self-efficacy is an important predictor of student achievement (Pajares, 1996; Klassen and Usher, 2010; Schunk and DiBenedetto, 2016). Students who have higher self-efficacy are more likely to persist on tasks, working through challenges they encounter to successfully complete the tasks (Bandura, 1986). Collaborative learning is one pedagogical technique that has been shown to positively correlate with student self-efficacy (Fencl and Scheel, 2005) and may thus benefit the development of self-efficacy in students. However, it is not clear how collaborative learning builds students' self-efficacy. In this study, we aimed to identify the specific experiences that occurred during collaborative group work that increased life science students' self-efficacy for quantitative biology problems. Additionally, we explored the extent to which students of different genders/sexes and with different initial self-efficacy levels relied on different experiences to increase their self-efficacy. We focused on quantitative tasks, because life science students report lower ability in mathematics than other science, technology, engineering, and mathematics (STEM) students (Sax *et al.*, 2015), which can affect their self-efficacy for quantitative tasks, and we were interested in identifying experiences that could be leveraged to support the development of self-efficacy for quantitative tasks in life science students. We found five different categories of experiences that increased students' self-efficacy for solving quantitative biology problems. In the following sections,

TABLE 4. Output for the generalized linear mixed models in which gender/sex or initial self-efficacy was significant: unstandardized regression coefficients (*B*), standard error of the regression coefficient (SE), *p* value, and odds ratio

	B	SE	*p* value	Odds ratio (e^B)
Accomplishing It: initial self-efficacy + gender/sex				
Intercept	−2.48	0.55	<0.001	0.08
Initial self-efficacy	0.39	0.12	0.001	1.48
Gender/sex: female	−0.15	0.26	0.56	0.86
Question type: PG	0.14	0.24	0.56	1.15
Course section: B	1.03	0.27	<0.001	2.81
Marginal R^2: 0.07				
Conditional R^2: 0.16				
Getting Help from Peers: gender/sex				
Intercept	−0.45	0.26	0.09	0.64
Gender/sex: female	0.62	0.28	0.03	1.87
Question type: PG	0.42	0.24	0.08	1.52
Course section: B	−0.86	0.29	0.003	0.42
Marginal R^2: 0.05				
Conditional R^2: 0.24				
Getting Help from Peers: initial self-efficacy + gender/sex				
Intercept	1.18	0.52	0.02	3.27
Initial self-efficacy	−0.45	0.13	<0.001	0.64
Gender/sex: female	0.47	0.28	0.10	1.59
Question type: PG	0.23	0.24	0.34	1.26
Course section: B	−0.97	0.29	<0.001	0.38
Marginal R^2: 0.09				
Conditional R^2: 0.26				

we elaborate on these experiences and how they relate to Bandura's (1997) sources of self-efficacy.

Simply Doing Problems Results in Mastery, but Peer Interactions Can Be Used to Gauge Success

In responses categorized as Accomplishing It, students described that the simple act of practicing, working through, or doing the HWE or PG problems increased their self-efficacy. These responses correspond to a mastery experience, which arises out of completing a task. They provide direct evidence to the students of their ability to succeed on a task and therefore are considered the most influential source of self-efficacy (Bandura, 1997). A mastery experience can be seen in this student's response: "I was able to easily calculate the genotype frequencies without any issues." This student uses the ease with which they calculated the answers as evidence of their mastery of the task. The relatively high frequency of responses coded as Accomplishing It (Figure 2) in this study supports the idea that

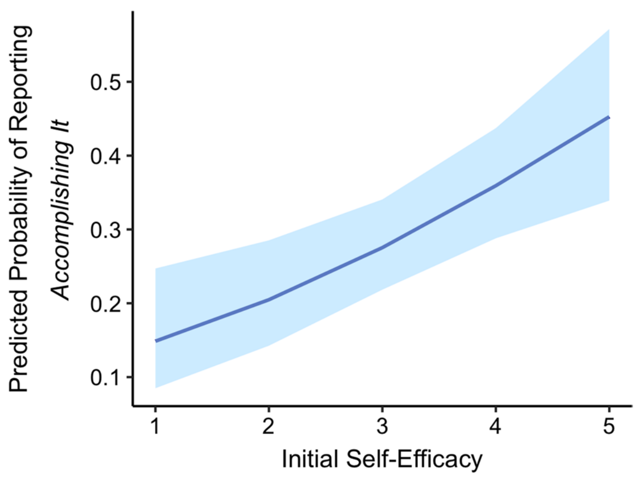

FIGURE 3. Relationship between initial self-efficacy and Accomplishing It. Data represent marginal means. The shaded area represents the 95% confidence interval.

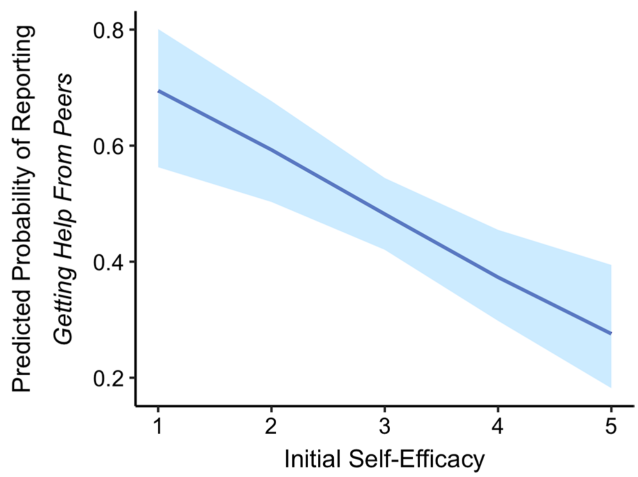

FIGURE 4. Relationship between initial self-efficacy and Getting Help from Peers. Data represent marginal means. The shaded area represents the 95% confidence interval.

mastery experiences are important for many students in the development of self-efficacy and aligns with the results of previous studies that have found mastery experiences to be a significant contributor to self-efficacy (Lent *et al.*, 1991; Britner and Pajares, 2006; Joët *et al.*, 2011; Byars-Winston *et al.*, 2017).

That said, mastery experiences through Accomplishing It do not rely on a group work setting. For example, one student reported, "It was the practice in general that helped me understand the concept, not practicing with others." This quote indicates that this type of mastery experience could be achieved outside a collaborative setting by having students complete practice problems on their own in class or as homework. However, our results suggest that mastery experiences can also be achieved through peer interactions during group work. Crucially, mastery experiences occur when students judge their efforts to be successful (Usher and Pajares, 2008). In this study, we found several codes that provide insight into how group work may afford unique opportunities for students to gauge their success on a task through peer interactions.

Explaining to a group member how to solve a part of the assignment or how to reason through a problem (Teaching/Guiding Others) enabled students to feel successful on the problem. This student described how teaching others reflected that they fully understood the material themselves, which increased their confidence in their ability to solve the problem on their own.

> I helped one of our group members who wasn't really getting it as much [to] solve the problems and understand which numbers to use where. This helped me by extension, as I know I am able to do these problems well enough as I can explain them to others.

Providing explanations to peers allows students to be metacognitive; they reflect on what they do and do not understand about a topic as they provide an explanation (Webb, 1989; Cooper, 1999). Thus, the experience of teaching others can serve as a metric by which students may gauge their success on a task, resulting in a mastery experience. However, teaching others may also increase a student's self-efficacy for a task through a vicarious experience. When teaching others, differences in knowledge between students become salient, and the student who is teaching others may interpret these differences in an evaluative way to inform a vicarious experience. In other words, students' self-efficacy may increase by virtue of perceiving that they are more capable than others in the class. Although we did not see evidence of any evaluative comparisons in our short-answer responses coded as Teaching/Guiding Others, vicarious experiences are notoriously difficult to measure (Usher and Pajares, 2008). We cannot rule out that these normative comparisons play a role in how students build their self-efficacy through teaching others, given how the sources of self-efficacy are often intertwined (e.g., Morris and Usher, 2011). However, these comparisons may not be salient enough, especially in comparison to the mastery experience, to be cited in a short-response survey. Regardless of which source of self-efficacy is tapped through teaching others, this experience may serve as an important practice for further bolstering self-efficacy in highly confident students. Students who are high achieving and have greater confidence in their abilities are more likely to explain concepts to other students in the group (Webb, 1989; Purzer, 2011). Thus, there is potential for collaborative group work to benefit not only students with low self-efficacy, but also those who already have high self-efficacy. Although we could not statistically test whether initial self-efficacy was related to reporting Teaching/Guiding Others, it is notable that the mean initial self-efficacy for those who reported this experience was 3.70 compared with 3.09 for students who did not report this experience (Table 3). Further research that includes a larger sample of students who are likely to report this experience would be useful for understanding which students may be more likely to rely on this experience to increase their self-efficacy for quantitative tasks.

Peer interactions in which students checked their answers with one another (Confirming Their Answers) were also integral for students in gauging their success on the problems. Webb-Williams (2018) also found that students' self-efficacy increased when they found they had the same answer as a peer; these were considered vicarious experiences, because students were using social comparisons to inform their self-efficacy beliefs. A social comparison is evident in this student's response: "I was able to do the problems and get the same answer as my group mates, which increased my confidence." However, we surmise that the vicarious experiences obtained through Confirming Their Answers may be entangled with mastery experiences in some cases. If students have done a problem, finding they have the same answer as other group members may suggest they have done the problem correctly, demonstrating mastery of the problem. The example quote used for Confirming Their Answers in the *Results*, "It helped reinforce that what I was doing was right," suggests the student is assessing their mastery of the problem. Another example of a student gauging their mastery can be seen in the quote below, where they explain that Confirming Their Answers meant they knew they were doing the problem correctly:

> I asked if for number 8 we should multiply the expected frequency of each genotype by the total number of individuals in the population and they said yes and we all got the same answers which reassured me that I knew what I was doing.

Thus, the experience of checking and confirming answers during group work is notable, because it provides opportunities for vicarious experiences to increase self-efficacy, through evaluating success based on peer comparisons, but may additionally serve as a mastery experience if students use answer-checking to judge their own abilities on a completed problem to be successful.

Help Availability During Group Work Builds Self-Efficacy

Being able to talk through ideas and receive help from peers (Getting Help from Peers) was also an important self-efficacy-building experience during group work for many students (Figure 2). Additionally, some students, though fewer in number, reported that help availability from an instructor or teaching assistant (Consulting with a Teacher) increased their self-efficacy. Although help availability is not one of Bandura's (1997) original four sources of self-efficacy, several qualitative studies have documented the value of help availability to students' self-efficacy, both in the form of peer support and instructor support (Hutchison *et al.*, 2006; Butz and Usher, 2015;

Usher *et al.*, 2019). Although getting help was specifically reported as a self-efficacy-building experience in our study, we surmise that help may relate to mastery and/or vicarious experiences. Additionally, help availability may also be related to social persuasions. Discussions among team members may involve encouragement, praise, or affirming messages as students work with one another to complete an assignment (Hutchison *et al.*, 2006; Purzer, 2011). In responses coded as Getting Help from Peers, students did not explicitly cite positive feedback as a mechanism by which working together improved their self-efficacy, but these affirming messages may have been too subtle for students to recall on a retrospective survey.

For Getting Help from Peers, a mastery experience may occur if help allows a student to successfully complete a problem. Asking questions or group discussion can help students to overcome obstacles they encounter on the problems, allowing them to be able to successfully master the problems. Although our short responses identify help availability as the source of increasing self-efficacy for quantitative problems, some of the responses reveal how a mastery experience may be intertwined with help availability. For example, this student reported that being able to discuss the problem with group members enabled her to understand how to do the problems: "Talking through each problem with group members helped me to better understand how to solve questions I wasn't confident on before." Another student explained how they mastered the problems on their own after being helped on the group work assignment: "My group members helped explain the parts of HWE that I didn't fully understand. They gave me a much more clear understanding of it, and now I can confidently get through it alone." Help that relates to a positive mastery experience is particularly interesting to consider, as help can sometimes dampen students' perceptions of mastery by diminishing their perceived personal involvement in their success (Bandura, 1997; Usher and Pajares, 2008). Future work is needed to understand whether a certain amount of help relates to a less powerful mastery experience, or whether particular students view the need for help as undermining, rather than improving, self-efficacy.

A vicarious experience can occur when Getting Help from Peers if help involves observing another student successfully working through the problem. The distinction between help availability relating to a mastery versus a vicarious experience is in whether help provides direct (mastery) or indirect (vicarious) evidence to the students of their capabilities. During group work, some group members may have sought help by watching another group member solve the problem. For example, this student explained how their confidence in their ability to do the problem increased as a result of help from other group members, which included observing a group member write out each step to solve the problem:

> As my group was solving the different problems, the steps were said out loud, discussed and we made sure everyone knew what was going on. It was good for me to watch someone solve the problem on paper step by step.

Thus, knowledgeable peers in the group have the potential to serve as models and initiate a vicarious experience when showing students how to solve a problem. However, the extent to which such help leads to a vicarious experience informing students' self-efficacy beliefs may depend on how similar the students receiving help perceives their knowledgeable peers to be to them.

Importantly, working with others to co-construct an answer may be another way to achieve a mastery experience. Discussion within a group can cue the prior knowledge of students, helping them to learn the material more deeply (Nokes-Malach *et al.*, 2015). Additionally, group members may have complementary knowledge of the topic (Nokes-Malach *et al.*, 2015), allowing students to rely on their peers to help them fill in their own knowledge gaps. This sentiment is expressed in this student response:

> My groupmates and I had a slight understanding of this concept, and once we shared what we knew, our group was able to piece together all the information we needed to know how to calculate the predicted number of individuals under H.W. equilibrium.

Dialogue in group work can help students co-construct knowledge, resulting in new ideas that neither student previously had (Chi and Wylie, 2014). This interactive mode of engagement, according to the interactive–constructive–active–passive (ICAP) framework, promotes deep learning (Chi and Wylie, 2014). We see deep learning as a possible mastery experience that engenders positive self-efficacy beliefs. Thus, promoting interactive dialogue in collaborative group work has the potential to reap significant benefits for students' self-efficacy.

The ability for peers to answer one another's questions and provide help is particularly critical in a large lecture setting with hundreds of students, where instructor attention is at a premium. The self-regulating and peer-guided nature of many groups enables them to resolve problems without necessarily relying on instructor oversight or intervention. Moreover, peers may be seen as less intimidating help providers than the instructor, as evidenced by this quote: "Talking through a roadblock in one of the questions was easier with my groupmates and not as uncomfortable as talking to a professor." Yet, in some cases, group members were unable to provide help to students, and help availability from the instructional staff became important for building self-efficacy. This can occur in situations where the entire group is confused, when the group's explanation fails to make sense to the student and they need an alternative explanation, or when dysfunctional group dynamics impede learning. The importance of getting help from an instructor can be seen in this student's response: "My group was confused on how to do one of the problems in the packet. When [the instructor] was walking around, [the instructor] helped us by explaining more in depth what to do. This helped us to understand what we were doing." This particular kind of experience highlights the importance of instructor interaction in a group work setting. Even though groups are able to self-guide and reinforce one another, not all issues can be overcome by the students or groups themselves. Having the fallback of reaching out to an instructor is still a significant component of improving students' self-efficacy.

Notably, we did not observe any responses that related to physiological and affective states. Students did not report that their self-efficacy increased as a result of positive feelings during the group work activity. However, this does not necessarily mean that physiological and affective states did not help to increase students' confidence in their ability to solve the quantitative tasks. Rather, it is possible that our survey was not set up to adequately capture this source of self-efficacy. Our survey asked students to describe "experiences and/or interactions," which may have prompted students to focus on external events rather than internal emotions. Researchers using similar prompts in other self-efficacy studies have discussed this as a possible explanation for the few responses that suggest physiological and affective states build self-efficacy (Butz and Usher, 2015; Usher *et al.*, 2019). It is also possible that physiological and affective states are more impactful in decreasing self-efficacy in this context than increasing self-efficacy. For example, Usher and colleagues (2019) found that none of the 173 coded responses indicated that a physiological and affective state increased a student's self-efficacy for science, and only 0.58% of the 334 coded responses indicated that a physiological and affective state increased a student's self-efficacy for math. In contrast, 2.37% and 6.74% of coded responses indicated that a physiological and affective state decreased a student's self-efficacy for science and math, respectively (Usher *et al.*, 2019). Negative feelings of stress, anxiety, and frustration during academic work may be particularly salient to students, impeding the development of confidence in their abilities to be successful in academic work. Further research is needed to better understand the role of physiological and affective states in building self-efficacy.

Given the group work setting of this study, the role of collective efficacy in building students' individual self-efficacy for quantitative tasks is also important to consider. Collective efficacy is a group's shared beliefs in their ability to successfully complete a task (Bandura, 1997). When tasks require interdependence, collective self-efficacy predicts group performance (Bandura, 1997; Katz-Navon and Erez, 2005). Several studies have shown that greater collective efficacy is associated with higher levels of group performance (Little and Madigan, 1997; Stajkovic *et al.*, 2009; Li *et al.*, 2015). Moreover, collective efficacy influences individuals' behaviors in a group through social norms and expectations (Goddard *et al.*, 2004). If a group has high collective efficacy, there are high expectations for success on the group task, and each group member may put in significant effort on the assignment as a result of these high expectations. In contrast, a group with low collective efficacy has low expectations for success on the group task, and thus it is unlikely group members will pressure one another to put in more effort and persevere when they encounter challenges (Goddard *et al.*, 2004). Individuals in groups with high collective self-efficacy, therefore, may have more opportunities to increase their individual self-efficacy through successfully completing the task, teaching others about the task when help is requested, or getting help from peers on the task when help is needed. Indeed, several studies have documented that collective self-efficacy predicts individual self-efficacy (Goddard and Goddard, 2001; Guidetti *et al.*, 2018; Skaalvik and Skaalvik, 2019). However, collective self-efficacy is understudied, especially in the context of student group work, and thus additional research is greatly needed to better understand how collective self-efficacy informs individual self-efficacy.

Together, our data illustrate how self-efficacy-building experiences during group work may draw on multiple sources of self-efficacy. The experiences that students have do not necessarily correspond to only one source of self-efficacy. Most of the experiences identified in this study may inform the development of students' self-efficacy through both mastery and vicarious experiences. Additionally, it is possible that social persuasions are part of these experiences, particularly when getting help from peers (e.g., Purzer, 2011). Crucially, our data cannot reveal whether students primarily interpret one source from an experience or whether they integrate multiple sources born from an experience. Mastery experiences are particularly influential in the development of self-efficacy (Bandura, 1997), and thus students may simply be interpreting different group work experiences as a mastery experience. On the other hand, individuals often integrate multiple sources of self-efficacy (Chen and Usher, 2013), and the combined effects of two or more sources from an experience may denote that as a particularly important self-efficacy-building experience for a student. Future studies could use interviews to more deeply probe these experiences and how they relate to the sources of self-efficacy. Interviews may also illuminate how context (e.g., students' perceptions of task difficulty or group members) plays a role in the group work experiences students report as building their self-efficacy for quantitative biology tasks.

Gender/Sex and Initial Self-Efficacy Relate to One Another and to Group Work Experiences That Build Self-Efficacy

We found that gender/sex did not predict the probability of students reporting that accomplishing the task increased their self-efficacy, but that female students had higher odds relative to male students of reporting that peer help increased their self-efficacy. These results align with previous work that found that mastery experiences are commonly cited sources of self-efficacy for both women and men (Butz and Usher, 2015; Usher *et al.*, 2019) and that socially oriented sources of self-efficacy may be more important for women than for men (e.g., Zeldin and Pajares, 2000; Zeldin *et al.*, 2008). However, consistent with other self-efficacy studies that examine gender (reviewed by Pajares, 2005), female students, on average, reported lower initial self-efficacy than male students. Therefore, we compared the output of a generalized linear model that contained gender/sex as the sole predictor of reporting an experience to the output of a model that contained both gender/sex and initial self-efficacy to explore whether differences in initial self-efficacy between genders/sexes might explain observed differences in the self-efficacy-building experiences reported by female students versus male students. When we entered initial self-efficacy into the Getting Help from Peers model, we found that students with lower initial self-efficacy had greater odds of reporting help availability from peers increased their self-efficacy, but gender/sex was no longer a significant variable. This suggests that initial self-efficacy may be mediating the relationship between gender/sex and the use of peer help to increase self-efficacy. In other words, female students may have greater odds than male students of reporting peer help availability because female students have lower initial self-efficacy than male students.

Interestingly, the relationship between gender/sex, initial self-efficacy, and the use of a particular group work experience to build self-efficacy is most evident in the less commonly reported group work experiences (Table 3). A high proportion of responses that indicated Teaching/Guiding Others increased self-efficacy were from male students, and the average initial self-efficacy scores of students who reported Teaching/Guiding Others was high (mean = 3.70). In contrast, a high proportion of responses that indicated Consulting with a Teacher increased self-efficacy were from female students, and the average initial self-efficacy scores of students who reported Consulting with a Teacher was low (mean = 2.18). However, caution should be used in drawing conclusions from these results, as only 27 responses reported Teaching/Guiding Others and only 34 responses reported Consulting with a Teacher. Given that only 32% of our entire sample was composed of male students, additional studies that incorporate larger sample sizes are needed to draw more robust conclusions about how gender relates to group work experiences that serve as sources of self-efficacy. In particular, we recommend that future studies be designed to test the extent to which factors, such as initial self-efficacy, mediate differences in the self-efficacy sources reported by men and women.

Students with higher initial self-efficacy had greater odds of reporting that completing the problems, which clearly represents a mastery experience, increased their self-efficacy, whereas students with lower initial self-efficacy had greater odds of reporting help availability from peers increased their self-efficacy. Our results corroborate previous findings in middle school students; students with high self-efficacy were more likely to report mastery experiences were influential (Butz and Usher, 2015). High self-efficacy scores reported by students may have at least somewhat accurately reflected their knowledge of and ability to solve these problems. Thus, they may have been well-positioned to complete the problems successfully without relying on help from other students. On the other hand, the students with low self-efficacy benefited from discussions and being able to ask questions about the problems to become more confident in their understanding of how to do them. Our results demonstrate that even students with high self-efficacy can benefit from a group work assignment. Moreover, these students may be important sources of self-efficacy, through help availability, for students with low initial self-efficacy, highlighting a key benefit of collaborative learning through group work. However, it is important to note that initial self-efficacy, in conjunction with the other fixed effects in the models, explained less than 10% of the variance in reporting an experience. Thus, other, unmeasured factors likely play an important role in determining the odds of a student reporting that completing the problems or peer help availability is a self-efficacy-building experience.

Limitations and Future Directions

The rating scale items that we designed and implemented limit the conclusions that we can draw. We created only one item to assess students' self-efficacy for an HWE or PG task. Therefore, we cannot make claims about students' self-efficacy for quantitative biology; we can only make claims about the self-efficacy for the specific tasks we asked students about: calculating the predicted number of individuals of each genotype under HWE conditions and predicting future population size using the logistic growth model. Additionally, our rating scale is limited in measuring change over time. Because we used only one item to measure self-efficacy, it is not possible to disentangle measurement error from the true change in score (Fuchs and Diamantopoulos, 2009). Moreover, response-shift bias, in which the internal standards on which a person bases a self-report change after an intervention, can obscure change results (Howard and Dailey, 1979). For example, pre-service teachers use a different internal metric to evaluate their self-efficacy after their first teaching experience; before teaching, they overestimate their teaching self-efficacy, because they do not know what they do not know (Cantrell, 2003; Cartwright and Atwood, 2014). Therefore, future studies should develop more robust, multi-item instruments and validate them as measures of self-efficacy for quantitative biology tasks in order to pursue quantitative studies that examine change in self-efficacy.

The open-response item was designed to capture and characterize a broad set of distinct group work experiences that increased students' self-efficacy for the HWE and PG tasks. However, we did not conduct cognitive interviews to understand the various ways in which students might interpret this question. Although we found that some students reported nothing from group work increased their confidence in their ability to solve the quantitative problems, we cannot rule out the possibility that some students may have felt compelled to write about a self-efficacy-building experience when in fact they did not believe that experience increased their self-efficacy. Additionally, the use of a short open-response question meant that, although we could survey a large number of students, we were unable to capture the range of students' experiences in great detail. In particular, recent research has shown that students may rely on several different sources of information or experiences to varying degrees when evaluating their self-efficacy beliefs (Chen and Usher, 2013), but our surveys likely only captured the most salient experiences through recall. Student responses typically only described one or two experiences and likely did not include the entire suite of experiences that they integrated to form their self-efficacy beliefs. Given the relatively greater influence of mastery experiences compared with other sources of self-efficacy (Bandura, 1997), it is likely that students easily recalled mastery experiences. However, observing similar peers succeed (vicarious experience), messages of encouragement (social persuasions), and positive feelings during an activity (physiological and affective states) are more subtle experiences and thus are less likely to be recalled, even if they did occur and impact students' self-efficacy. Moreover, given the subtlety of these experiences, students may not have been aware that these were the specific experiences that occurred when getting help from peers, for example, that increased their self-efficacy. To more thoroughly capture the full range of experiences that students use to inform their self-efficacy beliefs, future studies could interview students about their group work experiences. Students could be asked specific questions aimed at eliciting experiences that directly relate to each source of self-efficacy (e.g., Hutchison-Green *et al.*, 2008) and asked follow-up questions aimed at revealing the connections that may exist between multiple sources of self-efficacy. Additionally, while we were able to show a relationship between students' initial self-efficacy levels and their likelihood to report three common group work experiences, interviews could also provide more insight into how initial self-efficacy affects students' use of the less frequently reported experiences, such as Teaching/Guiding Others, in developing self-efficacy. Moreover, interviews would be useful for understanding how the specific characteristics of individual students, such as their academic background, socioeconomic status, and experiences relating to their race, ethnicity, and gender, influence their self-efficacy beliefs, how those beliefs are formed, and how they interpret group work experiences to develop those beliefs.

Additionally, our sample lacks demographic breadth to enable generalizations. The students surveyed were from two courses at one primarily White institution. Thus, we lack racial and ethnic diversity in our sample. Furthermore, our survey item conflated gender with sex, and the response choices contributed to othering students who do not identify as a male student or a female student. This may have resulted in underestimating the number of students with a queer gender identity. When "other" is used as a third response option on a gender item, fewer students report a queer gender than when a wide range of genders are included as response options (Casper *et al.*, 2022). Therefore, including an item with more-inclusive gender options, such as the item proposed by Casper and colleagues (2022), would allow us to better capture the experiences of students of all genders. Overall, collecting data from students from a wider variety of backgrounds and characteristics would allow for a better understanding of the variation in group work experiences used to build self-efficacy for quantitative biology tasks. Finally, the demographic makeup of groups should be considered in future studies, as that could influence the sources of self-efficacy that students use. For example, in group work situations where stereotype threat triggers negative perceptions of academic ability in minoritized students, vicarious experiences and social persuasions may become more important in forming self-efficacy beliefs than other sources (Usher and Pajares, 2008).

The context of how group work was implemented in the course sections could have affected our results, which also limits their generalizability. Students remained in groups for the entire semester, which may have affected any socially derived experiences that students reported as increasing their self-efficacy for the quantitative biology tasks. First, as students worked with one another over the course of the semester, their experiences may have changed as they became more comfortable with one another. At the beginning of the semester, students may have been more anxious as they worked to identify appropriate behaviors and roles, whereas at the end of the semester, students may have established group roles and norms and a sense of trust among group members (Wheelan, 2005). Therefore, getting help from peers, for example, may have become easier as the semester progressed. However, we did not observe a large difference in the frequency of codes reported at the end of the semester compared with the frequency of codes reported at the beginning of the semester (e.g., Discussing/Working Together only increased from 26% of responses on the HWE survey to 31% of responses at the time of the PG survey). That being said, a longitudinal analysis of experiences that students report as increasing their self-efficacy, in conjunction with data on how interactions among group members change over time, would provide a better understanding of how changes in group dynamics over time affect the sources of self-efficacy students tap. Second, because groups remained the same regardless of student attendance patterns, it is possible that not all of the group members were present on the group work days that were surveyed. Missing group members may change the dynamics of the group in a way that results in different experiences being reported. For example, the absence of a particularly collaborative group member may result in fewer opportunities for checking answers, peer help, or interactive dialogue that promotes mastery experiences. While understanding these group dynamics was not the focus of our study, it is important to consider the myriad ways in which group dynamics can change from class to class and how that influences the experiences students have during group work. Finally, the differential grading schemes of the two course sections could have affected the group work experiences that students in each course section reported. In course section B, where the group work assignment was graded for accuracy, students may have had less incentive to teach and guide others or ask questions, given that the stakes were higher to correctly complete the assignment in a given time period. Indeed, the results of the regression model indicate students in course section B had significantly lower odds of reporting a Getting Help from Peers experience than students in course section A (Table 4). Although this result may not be due specifically to the grading scheme in course section B, these data underscore the importance of understanding how contextual features of courses where group work is used affect students' self-efficacy-building experiences.

Going forward, there are several lines of research related to group work and self-efficacy that should be pursued. At the most basic level, it might be worthwhile to compare student self-efficacy after completing assignments as group work versus after completing assignments individually to identify the specific effect of group work on self-efficacy for different students. It would be particularly interesting to measure students' perceptions of mastery experiences in each condition to better understand the role of group work in engendering mastery experiences. Other avenues for future research include correlating self-efficacy-building group work experiences with achievement and investigating why self-efficacy-building group work experiences change between assignments. This latter research topic would provide insight into the observed differences in the group work experiences reported during the HWE group work versus the PG group work (Figure 2).

Finally, while we focused on the experiences during group work that increased students' self-efficacy, it is important to consider group work experiences that may have *decreased* students' self-efficacy. Group work is not without its limitations. Students often cite issues of unequal participation of group members and a perception of group activities as "busy work" (Chang and Brickman, 2018). Group dynamics can create stressful interactions between overconfident or overbearing students with their peers, stifling discussion and harming group cohesion, and ultimately impacting performance (Theobald *et al.*, 2017). Thus, group dynamics may break down and inhibit collaboration rather than foster it (Nokes-Malach *et al.*, 2015; Chang and Brickman, 2018; Donovan *et al.*, 2018). This can create experiences that may serve to decrease students' self-efficacy, such as preventing students from achieving a mastery experience, denying students peer help availability opportunities, or promoting negative social persuasions, which are theorized to be more impactful on self-efficacy than positive social persuasions (Bandura, 1997). Future work should focus on identifying students whose self-efficacy for a task decreased after group work and interviewing these students about the experiences that led to this decrease in self-efficacy.

Teaching Implications

This study explored how group work may influence self-efficacy in the context of quantitative biology, but our results have implications for understanding the impact of group work more broadly. There are several positive self-efficacy-building experiences that arise from group work that instructors can foster and that align with best practices for group work. Importantly, instructors should lean heavily into the collaborative benefits of group work and encourage students to discuss their ideas and results throughout the group work assignment, as this benefits students with lower self-efficacy, who in our study were more likely to be female students than male students. To facilitate collaborative discussion, instructors could assign group roles (e.g., Bailey *et al.*, 2012). Roles that can facilitate directed help, such as a "checker" who checks in on students' understandings of the procedures and concepts, may provide explicit opportunities for students to ask questions or be guided through a problem when they do not understand a procedure or concept. A role of "explainer" may encompass responsibilities of explaining and elaborating on concepts (Teaching/Guiding Others), benefiting the more confident students as well as their peers who need help. Reinforcing roles throughout the group work activity, and possibly even training students on roles, may be more effective than simply assigning roles (Chang and Brickman, 2018). Furthermore, assignments could be structured to include frequent checkpoints or opportunities for students to share and confirm their answers, which students in our study reported was beneficial. For example, instructors can incorporate problems into a group work assignment that ask students to discuss among themselves and form a consensus before proceeding or by segmenting a group work assignment for whole-class discussions (Gillies, 2013). These structures can help provide students with validation of their efforts and verification of their success and abilities throughout the group work process, increasing their self-efficacy. Finally, instructors should also check in with groups or monitor them frequently to ensure that students are able to work collaboratively and help one another and intervene when the self-guiding and self-checking apparatus of the group reaches a critical setback.

One challenge in implementing any pedagogical strategy is ensuring that as many students as possible receive the benefit of that strategy. A number of our students, particularly high self-efficacy students, reported that simply doing the problems increased their self-efficacy, suggesting working in groups is not necessary for these students. However, we envision several ways group work could be structured to still provide positive self-efficacy outcomes for these students. Instructors could allot some time at the beginning of the group work assignment for everyone to work independently, followed by collaborative group discussion. This would allow students the opportunity for a mastery experience through Accomplishing It, while also allowing for self-efficacy-building experiences through Getting Help from Peers. Alternatively, instructors could structure group work to engender mastery experiences through group work experiences other than Accomplishing It. For example, instructors could purposefully assign students with high self-efficacy as "explainers" during group work, giving them the opportunity for a mastery experience through Teaching/Guiding Others. Instructors should also consider designing challenging assignments that require students to work together to generate new ideas (Scager *et al.*, 2016). In such assignments, more students may rely on a mastery experience through peer help to increase their self-efficacy, given the difficulty in Accomplishing It on one's own. However, it is important for instructors to monitor student work during the early stages of the assignment and provide scaffolding as needed to ensure the difficulty of the assignment does not impede the development of students' self-efficacy (Lodewyk and Winne, 2005). To maximize group performance, on these challenging, interdependent assignments, and thus opportunities for positive self-efficacy-building experiences, it is important to consider collective efficacy. Maintaining the same groups for a long period of time and building in opportunities for groups to successfully complete assignments together (mastery experience) and receive positive feedback on their group performance (social persuasions) can help build collective efficacy (Baker, 2001; Katz-Navon and Erez, 2005). Importantly, when students are successful at more challenging tasks, it can potentially increase the significance and endurance of that success in shaping self-efficacy beliefs (Bandura, 1997; Usher and Pajares, 2008). However, it is not known whether substituting Accomplishing It experiences with other self-efficacy-building experiences will build self-efficacy in the same way. Future research should aim to determine the extent to which students rely on the specific experience of Accomplishing It to increase their self-efficacy versus other group work experiences that relate to mastery experiences.

Our results also suggest that grouping students heterogeneously with respect to self-efficacy levels has the potential to benefit both high and low self-efficacy students. We observed that low self-efficacy students benefited from peer help availability. Although we do not know who was giving help to these students in our study, Purzer (2011) found evidence that students with higher initial self-efficacy were more likely to tutor others and provide explanations in a team setting. In our study, teaching others was an important way for students to build their self-efficacy through group work. Thus, promoting question asking and teaching/guiding experiences in heterogeneous groups can simultaneously improve self-efficacy in both low and high self-efficacy students, respectively. Indeed, research examining group composition by ability has suggested that low- and high-ability students can both benefit from heterogeneous groups in this way (Webb, 1982; Lou *et al.*, 2001, 1996). However, it is important to note that these studies also demonstrated that medium-ability students benefit from homogeneous groups, and other studies have demonstrated that low- or high-ability students have greater cognitive outcomes in homogeneous groups (e.g., Baer, 2003; Jensen and Lawson, 2011). Although self-efficacy and academic ability are different, self-efficacy is a strong predictor of academic performance (Schunk and DiBenedetto, 2016). Thus, grouping by self-efficacy scores may produce similar groups to those that would be created by grouping by measures of ability, such as pretest scores or GPA. Given the mixed results on the effects of homogeneous and heterogeneous groups on student outcomes, further studies should seek to understand under what conditions—and for whom—heterogeneous self-efficacy groups would be beneficial.

CONCLUSION

We found that introductory biology students working in groups to complete quantitative biology tasks such as calculating HWE and modeling population growth build their self-efficacy from a variety of group work experiences. Successfully completing practice problems, a mastery experience, was one of the most frequently reported experiences that increased students' self-efficacy. Although completing practice problems is not unique to a group work setting, as students may be tasked with individual practice through homework assignments, we found that students also identified a number of self-efficacy-building experiences that relied on interacting with others in the group or in the classroom. These experiences include checking answers with group members, teaching group members, receiving help from group members, and receiving help from the instructor. These experiences have the potential to encompass several sources of self-efficacy, including mastery experiences, vicarious experiences, and/or social persuasions. Moreover, students may interpret multiple sources from an experience given the interconnected nature of the sources of self-efficacy. Regardless of which source of self-efficacy underlies an experience, peer interactions in group work can provide instantaneous or immediate benefits to a student's self-efficacy over the course of the assignment.

Many contextual and personal factors govern which sources a student weighs in the development of self-efficacy. Our data demonstrate that students' pre-existing self-efficacy beliefs play a role in the interpretation of self-efficacy-building experiences. Help availability is particularly important to students with lower initial self-efficacy, who are more likely to be women, whereas simply doing the practice problems is more likely to build self-efficacy in students with higher initial self-efficacy.

Overall, our results add to research documenting the benefits of group work by identifying specific experiences that are important for building students' self-efficacy. We have discussed how the specific experiences may relate to the sources of self-efficacy, providing a foundation for future research to explore the process more deeply. From a teaching perspective, our results highlight the importance of providing students with opportunities to demonstrate their mastery, while also building in frequent discussion questions or checkpoints to reinforce and encourage group members to collaborate with one another. Instructors seeking to build life science students' quantitative skills may find collaborative group work to be a particularly effective strategy. In particular, those students who are not confident in their abilities to be successful on quantitative tasks may find the help that peers provide to be important for improving their self-efficacy beliefs.

ACKNOWLEDGMENTS

We would like to thank Jacqueline Klombers and Ciara Devereaux for assisting in data analysis. We would also like to thank the instructor of the other course section for allowing us to survey the students. Finally, we would like to thank several anonymous reviewers for their feedback which has considerably improved the article.

REFERENCES

Ahn, H. S., Usher, E. L., Butz, A., & Bong, M. (2016). Cultural differences in the understanding of modelling and feedback as sources of self-efficacy information. *British Journal of Education Psychology*, *86*, 112–136.

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.

Association of American Medical College-Howard Hughes Medical Institute. (2009). Washington, DC: *Scientific foundations for future physicians*.

Baer, J. (2003). Grouping and achievement in cooperative learning. *College Teaching*, *51*(4), 169–174.

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.

Baker, D. F. (2001). The development of collective efficacy in small task groups. *Small Group Research*, *32*(4), 451–474.

Baldwin, J. A., Ebert-May, D., & Burns, D. J. (1999). The development of a college biology self-efficacy instrument for nonmajors. *Science Education*, *83*(4), 397–408.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*, ar56.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In Pajares, F., & Urdan, T. C. (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586–598.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485–499.

Butz, A. R., & Usher, E. L. (2015). Salient sources of early adolescents' self-efficacy in two domains. *Contemporary Educational Psychology*, *42*, 49–61.

Byars-Winston, A., Diestelmann, J., Savoy, J. N., & Hoyt, W. T. (2017). Unique effects and moderators of effects of sources on self-efficacy: A model-based meta-analysis. *Journal of Counseling Psychology*, *64*(6), 645–658.

Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: A multiple-groups analysis. *Journal of Counseling Psychology*, *57*(2), 205–218.

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis) measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, *13*(11), e12506.

Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics*, *103*(4), 177–185.

Cartwright, T. J., & Atwood, J. (2014). Elementary pre-service teachers' response-shift bias: Self-efficacy and attitudes toward science. *International Journal of Science Education*, *36*(14), 2421–2437.

Casper, A. A., Atadero, R. A., & Fuselier, L. C. (2022). Revealing the queer-spectrum in STEM through robust demographic data collection in undergraduate engineering and computer science courses at four institutions. *PLoS ONE*, *17*(3), e0264267.

Chang, Y., & Brickman, P. (2018). When group work doesn't work: Insights from students. *CBE—Life Sciences Education*, *17*(3), ar52.

Chen, J. A., & Usher, E. L. (2013). Profiles of the sources of science self-efficacy. *Learning and Individual Differences*, *24*, 11–21.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Cooper, M. A. (1999). Classroom choices from a cognitive perspective on peer learning. In O'Donnell, A. M., & King, A. (Eds.), *Cognitive perspectives on peer learning* (pp. 215–234). Mahwah, NJ: Erlbaum.

Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE—Life Sciences Education*, *17*(4), ar60.

Drott, M. T., & Sarvary, M. A. (2016). *Why did the snake cross the road? A population genetics and habitat conversation case study*. Arlington, VA: National Center for Case Study Teaching in Science. Retrieved April 4, 2022, from www.nsta.org/ncss-case-study/why-did-snake-cross-road

Fencl, H., & Scheel, K. (2005). Engaging students: An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course. *Journal of College Science Teaching*, *35*(1), 20–24

French, L. R., Walker, C. L., & Shore, B. M. (2011). Do gifted students really prefer to work alone? *Roeper Review*, *33*, 145–159.

Franceschini, G., Galli, S., Chiesi, F., & Primi, C. (2014). Implicit gender–math stereotype and women's susceptibility to stereotype threat and stereotype lift. *Learning and Individual Differences*, *32*, 273–277.

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195–210.

Gillies, R. M. (2013). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, *39*, 35–49.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science Motivation Questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176.

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education*, *17*(7), 807–818.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, *33*(3), 3–13.

Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: A problem hiding in plain sight. *BMJ*, *353*, i1981.

Guidetti, G., Viotti, S., Bruno, A., & Converso, D. (2018). Teachers' work ability: A study of relationships between collective efficacy and self-efficacy beliefs. *Psychology Research and Behavior Management*, *11*, 197–206.

Holley, J. W., & Guilford, J. P. (1964). A note on the G index of agreement. *Educational and Psychological Measurement*, *24*(4), 749–753.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, *64*(2), 144–150.

Hruschka, D. J., Schwartz, D., John, St., D., C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, *16*(3), 307–331.

Hutchison, M. A., Follman, D. K., Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, *95*(1), 39–47.

Hutchison-Green, M. A., Follman, D. K., & Bodner, G. M. (2008). Providing a voice: Qualitative investigation of the impact of a first-year engineering experience on students' efficacy beliefs. *Journal of Engineering Education*, *97*(2), 177–190.

Jensen, J. L., & Lawson, A. (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE—Life Sciences Education*, *10*(1), 64–73.

Joët, G., Usher, E. L., & Bressoux, P. (2011). Sources of self-efficacy: An investigation of elementary school students in France. *Journal of Educational Psychology*, *103*(3), 649–663.

Jones, B. D., Paretti, M. C., Hein, S. F., & Knott, T. W. (2010). An analysis of motivation constructs with first-year engineering students: Relationships among expectancies, values, achievement, and career plans. *Journal of Engineering Education*, *99*(4), 319–336.

Kassambara, A. (2022). *rstatix: Pipe-friendly framework for basic statistical tests*. R package version 0.7.1. Retrieved January 9, 2023, from https://CRAN.R-project.org/package=rstatix

Katz-Navon, T. Y., & Erez, M. (2005). When collective- and self-efficacy affect team performance: The role of task interdependence. *Small Group Research*, *36*(4), 437–465.

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In Karabenick, S. & Urdan, T. C. (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1–33). Bingley, UK: Emerald Group.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, *33*(3), 265–269.

Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of Counseling Psychology*, *38*(4), 424–430.

Li, X., Zhou, M., Zhao, N., Zhang, S., & Zhang, J. (2015). Collective-efficacy as a mediator of the relationship of leaders' personality traits and team performance: A cross-level analysis. *International Journal of Psychology*, *50*(3), 223–231.

Little, B. L., & Madigan, R. M. (1997). The relationship between collective efficacy and performance in manufacturing work teams. *Small Group Research*, *28*(4), 517–534.

Lodewyk, K. R., & Winne, P. H. (2005). Relations among the structure of learning tasks, achievement, and changes in self-efficacy in secondary students. *Journal of Educational Psychology*, *97*(1), 3–12.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*(3), 449–521.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423–458.

Matsui, T., Matsui, K., & Ohnishi, R. (1990). Mechanisms underlying math self-efficacy learning of college students. *Journal of Vocational Behavior*, *37*, 225–238.

Morris, D. B., & Usher, E. L. (2011). Developing teaching self-efficacy in research institutions: A study of award-winning professors. *Contemporary Education Psychology*, *36*(3), 232–245.

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213.

National Research Council (NRC). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.

NRC. (2009). *A new biology for the 21st century*. Washington, DC: National Academies Press.

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*(4), 645–656.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. In Gallagher, A. M., & Kaufman, J. C. (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 294–315). Cambridge, UK: Cambridge University Press.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*(2), 124–139.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Purzer, S. (2011). The relationship between team discourse, self-efficacy, and individual achievement: A sequential mixed-methods study. *Journal of Engineering Education*, *100*(4), 655–679.

R Core Team. (2021). *R: A language and environment for statistical computing, version 4.1.2*. Vienna: R Foundation for Statistical Computing. Retrieved April 4, 2022, from www.R-project.org

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Washington, DC: Sage.

Sawtelle, V., Brewe, E., Goertzen, R. M., & Kramer, L. H. (2012). Identifying events that impact self-efficacy in physics learning. *Physical Review Special Topics—Physics Education Research*, *8*(2), 020111.

Sax, L. J., Kanny, M. A., Riggers-Piehl, T. A., Whang, H., & Paulson, L. N. (2015). "But I'm not good at math": The changing salience of mathematical self-concept in shaping women's and men's STEM aspirations. *Research in Higher Education*, *56*, 813–842.

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.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, *50*(11), 835–850.

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In Wentzel, K. R., & Miele, D. B. (Eds.), *Handbook of motivation at school* (pp. 34–54). New York, NY: Routledge.

Sheu, H-B., Lent, R. W., Miller, M. J., Penn, L. T., Cusick, M. E., & Truong, N. N. (2018). Sources of self-efficacy and outcome expectations in science, technology, engineering, and mathematics domains: A meta-analysis. *Journal of Vocational Behavior*, *109*, 118–136.

Skaalvik, E. M., & Skaalvik, S. (2019). Teacher self-efficacy and collective teacher efficacy: Relations with perceived job resources and job demands, feeling of belonging, and teacher engagement. *Creative Education*, *10*, 1400–1424.

Smith, B. L., & MacGregor, J. T. (1992). What is collaborative learning? In Goodsell, A. S., Maher, M. R., Tinto, V., Smith, B. L., & MacGregor, J. (Eds.), *Collaborative learning: A sourcebook for higher education* (pp. 10–29). University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Stajkovic, A. D., Lee, D., & Nyberg, A. J. (2009). Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model. *Journal of Applied Psychology*, *94*(3), 814–828.

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

Usher, E. L., Ford, C. J., Li, C. R., & Weidner, B. L. (2019). Sources of math and science self-efficacy in rural Appalachia: A convergent mixed methods study. *Contemporary Educational Psychology*, *57*, 32–53.

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*(4), 751–796.

Warfa, A.-R. M. (2016). Mixed-methods design in biology education research: Approach and uses. *CBE—Life Sciences Education*, *15*, rm5.

Webb, N. M. (1982). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology*, *74*(5), 642–655.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*(1), 21–39.

Webb-Williams, J. (2018). Science self-efficacy in the primary classroom: Using mixed methods to investigate sources of self-efficacy. *Research in Science Education*, *48*, 939–961.

Wheelan, S. A. (2005). The developmental perspective. In Wheelan, S. A. (Ed.), *The handbook of group research and practice* (pp. 119–132). Thousand Oaks, CA: Sage.

Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology*, *82*(6), 1219–1227.

Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching*, *45*(9), 1036–1058.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal*, *37*(1), 215–246.

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education*, *25*(9), 1081–1094.