Article

Caution, Student Experience May Vary: Social Identities Impact a Student's Experience in Peer Discussions

Sarah L. Eddy,*† Sara E. Brownell,†‡ Phonraphee Thummaphan,§ Ming-Chih Lan,§ and Mary Pat Wenderoth‖

*College of Natural Sciences, University of Texas at Austin, Austin, TX 78712; ‡School of Life Sciences, Arizona State University, Tempe, AZ 85281; §College of Education and ‖Department of Biology, University of Washington, Seattle, WA 98195

In response to calls for implementing active learning in college-level science, technology, engineering, and mathematics courses, classrooms across the country are being transformed from instructor centered to student centered. In these active-learning classrooms, the dynamics among students becomes increasingly important for understanding student experiences. In this study, we focus on the role a student prefers to assume during peer discussions, and how this preferred role may vary given a student's social identities. In addition we explore whether three hypothesized barriers to participation may help explain participation difference in the classroom. These barriers are 1) students are excluded from the discussion by actions of their groupmates; 2) students are anxious about participating in peer discussion; and 3) students do not see value in peer discussions. Our results indicate that self-reported preferred roles in peer discussions can be predicted by student gender, race/ethnicity, and nationality. In addition, we found evidence for all three barriers, although some barriers were more salient for certain students than others. We encourage instructors to consider structuring their in-class activities in ways that promote equity, which may require more purposeful attention to alleviating the current differential student experiences with peer discussions.

INTRODUCTION

Active learning has been shown to improve the academic performance of undergraduates in science, technology, engineering, and mathematics (STEM) courses (Freeman *et al.*, 2014), and these findings have motivated national efforts to transform STEM classrooms into interactive, student-centered learning spaces (American Association

for the Advancement of Science, 2011; Singer *et al.*, 2013). As we make this shift in pedagogy, we need to consider not only the dynamics between the instructor and student, but also the dynamics among the students themselves. In this study, we explore how social identities (e.g., gender, race/ethnicity, nationality) could influence a student's experience in one of the most common implementations of active learning: peer discussions (Crouch and Mazur, 2001; Borrego *et al.*, 2013).

Peer discussions, as we operationally define them in this paper, involve small groups of students talking with one another in class about course content. Peer discussions are a component of many types of classroom activities, including short activities such as answering clicker questions or longer activities such as case studies for which students may work in groups for an entire class period. Participation during peer discussions is generally voluntary, as the size of most STEM classrooms prohibits instructors from enforcing an individual's level of engagement. These peer discussions are often accompanied by some kind of debriefing that can involve students reporting out to the whole class what their groups discussed.

In general, peer discussions in undergraduate courses seem to increase student learning compared with courses without peer discussions (Johnson and Lawson, 1998; Springer *et al.*, 1999; Crouch and Mazur, 2001; Smith *et al.*, 2009). However, some studies have revealed a need for caution when interpreting these results, because some students may be benefiting more than others from these interactions. While some of the differences in learning between students in these studies can be attributed to the expected noise associated with education studies, there is some evidence of a systematic bias in who benefits most from peer discussions. For example, Beichner *et al.* (2007) found the top third of the class benefited the most from peer discussions, and Jensen and Lawson (2011) observed a similar pattern: when below-average students were paired with higher-achieving students for peer discussions, the higher-achieving students demonstrated greater learning gains than the other students in the group.

The prevailing hypothesis for these performance differences is that some students did more of the explaining than other students (Beichner *et al.*, 2007; Jensen and Lawson, 2011). Although this hypothesis has not been directly tested in college classrooms, it has theoretical and observational support. Theoretical support comes from the ICAP (Interactive > Constructive > Active > Passive) framework, which hierarchically categorizes the different ways that students can engage in a classroom (Chi and Wylie, 2014). This framework connects the level of student engagement with how students encode new material, which in turn affects learning outcomes. Students who are silent and only listen to their groupmates during peer discussions engage either passively (e.g., listening) or actively (e.g., writing notes). These forms of engagement are predicted to help students store but not integrate new information. Students who engage in explaining (e.g., talking) during group work are engaging either in constructive engagement, in which the student constructs his or her own mental model, or interactive engagement, in which the student builds on another student's ideas. Both of these forms of engagement lead to the integration of new ideas with a student's prior knowledge and schemas. The more integrated the learning, the more likely the student is to develop a deep understanding of the material, whereas students who merely listen during group work seem to gain only a shallow understanding (Chi and Wylie, 2014). The difference between superficial and deep learning could account for observed differences in learning gains in active-learning STEM classrooms. This differential learning in peer discussions has been demonstrated with psychology students in a laboratory setting: students who provided an explanation aloud performed better on a knowledge assessment than students who listened to and critiqued in writing the explanation they heard (Willoughby *et al.*, 2000). Thus, there is theoretical support for the assertion that the role a student plays in peer discussions (e.g., explaining or listening) may explain differences in learning outcomes.

At the most basic level, before we can argue that the different roles students play in peer discussions could explain why we see variation in learning outcomes, we have to demonstrate that different students play different roles in these discussions. Observationally, there is extensive evidence for the existence of inequalities in who assumes the explaining role in peer discussions in STEM classrooms. For example, peer discussions around clicker questions in physics and astronomy classrooms are frequently dominated by a single student, particularly when students get credit only for correct answers (James *et al.*, 2008; James and Willoughby, 2011). By definition, if one student dominates, then the other students in the group do not explain as much during the discussion and have less opportunity to engage interactively or constructively. Knight *et al.* (2013) found similar differential participation in biology classrooms, with only one student explaining his or her reasoning in ~20% of the recorded peer discussions. Even asynchronous online group discussions in an engineering course frequently revealed a dominator (Ocker, 2007). Thus, it seems inequities in participation are present in STEM classrooms, and this provides support for the idea that differences in the roles that students play in peer discussions have the potential to impact learning.

Previous studies that explore factors that influence the roles that students play in peer discussions in college STEM classrooms have focused on how students with different levels of competency with course material differ in the degree to which they assume explaining roles in peer discussions (Beichner *et al.*, 2007; Jensen and Lawson, 2011), yet competency in a discipline is not the only factor that may be important. Social identities, our sense of who we are based on our group memberships, may also influence how a student participates in peer discussions.[1] For example, gender has been shown to impact whole-class participation; women answer instructor-posed questions less than men in both traditional face-to-face STEM classrooms and online classrooms (Eddy *et al.*, 2014; Koenig, 2015). Additionally, race/ethnicity has been shown to impact participation in college classrooms (Howard *et al.*, 2006; White, 2011), although we are only aware of one study that has examined these differences specifically in STEM classrooms (Eddy and Hogan, 2014). Eddy and Hogan found that black students in a traditional biology classroom at an R1 university were 2.3 times more likely to report lower in-class participation relative to white students, although this difference dissipated when a higher level of active learning was implemented. If social identities influence participation in whole-class discussions, social identities could also contribute to differences in how students participate in peer discussions.

In this paper we identify differences in how students of different races/ethnicities, nationalities (international vs. American national), and genders engage in peer discussions using student self-report data from three large introductory biology classrooms. Specifically, in Study I, we explore whether race/ethnicity, nationality, and gender predict the roles that students prefer to play in peer discussions (e.g., explaining or listening), as these roles likely impact a student's opportunity to learn from peer discussion.

[1]It is important to note that students hold multiple social identities (including gender, race/ethnicity, sexual orientation, generation status, ability status), although some of these may be more salient to them than others. Nevertheless, the unique combination of identities that students hold likely influences their experiences in college classrooms. The impacts of this intersectionality of identities has not been well explored in college-level biology; most studies only focus on the impact of one or two identities on student experience. For an example of a paper exploring intersectionality in STEM classrooms, see Ong (2005).

We argue that preferred roles is one possible proxy for the role students actually play, on average, in these discussions, given the freedom to change groups that unstructured group work provides. This freedom allows students to seek out groups in which they are most comfortable, which will likely be where, on average, they can assume their preferred roles. We recognize, however, that some classroom settings may not encourage students to change groups through the term and that even when students do change groups, they might not do so in a way that allows them to optimally select their preferred roles.

Identifying inequities in the roles that students assume is important, but it does not provide instructors with guidance on how to facilitate more equitable participation. To take effective action, an instructor has to be aware of which barriers contribute to differential participation patterns across social identities. These barriers can include factors external to a student, such as how they are treated by those around them (Morris, 2007) and whether or not they are called on by the instructor. Barriers can also be factors originating from within students themselves, such as their own comfort speaking in front of people or whether they understand a question well enough to contribute to the discussion. One strategy for identifying potential barriers to participation is to query students about their experiences with these barriers. For example, if international students report more anxiety about participation than American students and also show lower likelihood to engage in explaining in their groups than American students, then it would seem that anxiety might be influencing how they engage in peer discussions. If international students are more comfortable or equally comfortable relative to American students, then we would lack evidence to support anxiety as a potential barrier for this group of students. By identifying barriers to equal participation, researchers and instructors could develop targeted interventions to encourage the participation from groups of students who tend to participate less in unstructured group settings.

In Studies II–IV we took the first exploratory steps toward identifying possible barriers to participation in peer discussions using student self-report on the degree to which students experience barriers. We focused on three possible barriers (one external and two internal): 1) students are excluded from the discussion by actions of their groupmates; 2) students are anxious and uncomfortable with participation; and 3) students do not see value in peer discussions. To evaluate each factor's ability to explain the differences in preferred roles reported by specific student groups, we compare the frequency with which different groups of students report these barriers with the patterns in self-reported preferred roles. These three barriers are not the only possible barriers influencing student participation but serve as a starting point to inform future investigations of issues of equity in peer discussions.

Hypothesized Barrier 1 (External): Students Are Excluded from Peer Discussion by Groupmates

The ability to participate in peer discussions can be limited by external factors, such as the presence of a dominating student in peer discussion (James *et al.*, 2008). Few studies at the college level have disaggregated student data to identify who reports being dominated and who does not (for an exception, see Kelsey, 2000). Other ways that groupmates could limit participation include when the group moves too quickly through the material for a student to keep up and contribute (Kotsopoulos, 2010) or when a student's contributions are demeaned or ignored by other students so that a student becomes reluctant to offer ideas (White, 2011). In Study II, we explore whether students of different social identities differ in their perceptions that groupmates are a barrier to their own participation.

Hypothesized Barrier 2 (Internal): Student Anxiety about Participation Limits Their Participation

The sense of anxiety and intimidation that students feel in the classroom can affect their willingness to participate (Karp and Yoels, 1976; Peters, 1978; Fritschner, 2000; Hyde and Ruth, 2002; Weaver and Qi, 2005; Micari and Drane, 2011). In Study III, we explore whether students of different social identities differ in the level of their concerns about participation in peer and whole-class discussions in a biology classroom.

Hypothesized Barrier 3 (Internal): Low Student Perceptions of the Value of Peer Discussions Limit Their Participation

If students see value in completing the activity, then they are more likely to engage in that activity (Meece *et al.*, 2006). Value can be extrinsic (e.g., if an instructor assigns points for completing an assignment, then a student may complete it to get a good grade). However, when a task is voluntary, such as engaging in peer discussions, the value must be intrinsic. This motivation can arise if students believe that their participation in the discussion will result in the attainment of their goals, such as learning from the activity (Oliver, 1974). There is some evidence that cultural or socialization differences around the importance of talking could lead to differences in students' perceptions of the value of peer discussions (Kim, 2002, 2008). In Study IV, we explore whether students' social identities predict differences in student perceptions of the value of peer discussions.

In summary, while much literature exists on the aggregate learning gains from active learning (Freeman *et al.*, 2014) or how students in aggregate feel about active learning (Johnson and Lawson, 1998; Springer *et al.*, 1999), there is a lack of knowledge about *how different groups of students* engage with active learning and what factors might be contributing to those differences. Overall, in this study, we explore the ways that different groups of students prefer to engage in one of the most common methods of implementing active learning: peer discussions. We assess the presence and frequency of three barriers that we predict could explain these differences, and we use our results to create a more holistic picture of why experiences with peer discussion might differ among students of different races/ethnicities, nationalities, and genders. As more classrooms transition to a student-centered approach, having a better sense of the barriers to participation could enable us to better structure peer discussion to maximize learning for all students.

METHODS AND RESULTS

The Classes and Student Population

This study spans three terms of one introductory biology course for majors. This course was the second in a three-course sequence and focuses on molecular, cellular, and developmental biology. Class sizes ranged from less than 100 to more than 700 students (term A, $n = 95$; term B, $n = 260$; term C, $n = 705$). All three iterations of this course were taught in an intensely active-learning way, with daily small peer discussions (two to four students) as well as whole-class discussions with students reporting out in front of the entire class. In addition, longer activities that encompassed almost the entire class period were implemented weekly. These activities involved students working on a worksheet that engaged them in a series of questions and discussions related to biology content. Students worked in self-selected groups based on where they chose to sit in the large classroom; they were not required to sit in any particular section of the classroom. The surveys administered in this study were completed online the night after they completed one of the worksheet activities in class.

Students enrolled in these courses were primarily sophomores (48.9%) and juniors (40.3%). Female[2] students made up 58.9% of the classroom population. In addition, of the students enrolled in the course, 44.2% were Asian Americans,[3] 39.5% were white Americans, 6.3% were international, 5.5% were Latin@ Americans, 1.8% were black Americans, 1.8% were Hawaiian and Pacific Islander, and 0.8% were Native Americans. The relatively low number of Latin@, black, Hawaiian and Pacific Islander, and Native American students made it necessary to pool their responses into one category for the analyses presented in this paper, despite our recognition that these groups can face different barriers in the classroom and respond differently to treatments (cf. Eddy and Hogan, 2014). These populations are united by the shared experience of being both historically and currently underserved by institutions of education in America (Ladson-Billings, 2006) and, thus, we use the title underserved students to refer to these populations of students.

Differences Between Students

Students vary in many ways that could account for differences in their feelings about group participation. We hypothesized that student level factors such as gender (a factor with two levels: female and male) and race/ethnicity/nationality (a factor with four levels: Asian American, underserved American, white American, and international students) could influence student responses. We collected demographic data from the registrar to determine gender and race/ethnicity/nationality. To distinguish differences based on identity from differences based on competency with the course material, we included total exam points (a continuous variable) in the course in which the survey was taken as a point estimate of a student's demonstrated competency as a control in all the analyses presented in this paper. Exam points earned was converted into a z-score ((student exam points earned – mean student exam points earned in course)/SD in exam points earned). We are calling this measure "relative biology competency," as it is relative to the other students in that course. At the time students took the surveys used in this study, they had already taken three of four exams. We found that student performance on each individual exam was very strongly and positively correlated with total exam points earned as well as performance on the final in all three terms, so we felt confident in using total exam points as our proxy for relative biology competency (Supplemental Material Table 1). In addition, we had information on the grades students earned in a prerequisite biology course and their cumulative college GPAs at the start of the current biology class. Both of these external measures of aptitude were strongly and positively correlated with student relative biology competency in the current course.

Differences Between Peer Discussion Groups

The particular combination of students in a group could also influence a student's experience and perception of peer discussions. Students formed their own groups, so we could not control who worked with whom or size of the groups (although groups ranged in size from two to four). We did collect data on whether a student had a friend in a group using a four-level Likert-scale question. The responses on this question were collapsed to a binary, with "strongly disagree" and "disagree" indicating no friend in the group and "strongly agree" and "agree" indicating a friend in the group. In analyses in which it was possible, we also explored an interaction between gender and friend in group and race/ethnicity/nationality and friend in group. We chose to explore these two interactions, because prior research had established that working with a friend influenced males and females differently in group work (Strough *et al.*, 2001). More generally, research on students in domains in which they are stereotyped has demonstrated the importance of friends in the domain for students from historically underrepresented groups such as black students relative to white students (Walton and Cohen, 2007) and women relative to men (Robnett, 2013).

---

[2]It is well established that gender does not exist as a binary variable but rather as a spectrum of identities (for an introductory-level primer, see Stryker, 2008). Unfortunately, many universities, including the one in this study, continue to include only male and female options on their applications. Thus, for our study, we are limited to reporting only male and female genders, which may mask other identities.

[3]It is also important to note that pooling Asian-American students into one category could be equally as problematic as grouping students into the underserved category. Asian-American students come from a variety of backgrounds and cultural contexts that also could lead to them experiencing different barriers in the classroom. For example, at the institution in this study (as of 2011), a wide variety of cultural backgrounds were represented in the Asian-American sample: 19.8% of Asian-American students were Chinese Americans, 16.0% were Korean, 11.1% were Vietnamese, 9.4% were Filipino, 6.7% were Asian Indian, and 33.2% reported other Asian backgrounds. Unfortunately, at the time of the study we were unable to obtain more detailed ethnicities for the students in our specific classrooms. Thus, we use "Asian-American students" as a category that captures barriers and challenges in the classroom shared across all these ethnicities of students. As with underserved students, this pooled approach may mask barriers faced by individual groups of students.

Controlling for Differences between Terms

In all our models, we include a variable for the term that students were enrolled in the class. This variable is almost always significant and indicates that the overall magnitude of the effect varies by term. Importantly, we never observe an interaction between term and any of our variables of interest (e.g., gender). This indicates that the patterns that we see are consistent across terms, but the overall scale may vary, potentially due to differences in the instructor, size, or student composition of particular classes. Because the pattern is consistent, we do not discuss the impacts of term in the Results section, although the regression coefficient is always present in the models. The regression coefficients can be found in the Supplemental Materials.

General Statistical Approach: Model Selection and Multimodel Inference

Overview. The studies in this paper are exploratory. We have identified multiple factors that differ among students (gender, race/ethnicity/nationality) that may lead to differences in their experiences associated with peer discussions. We do not have a priori hypotheses about which of these variables will be important in which context. Thus, null-hypothesis testing is not appropriate, and we instead use model selection and multimodel inference as our statistical approach (Burnham and Anderson, 2003). This approach starts with a full model including all possible reasonable relationships between the predictor variables and then, using model-selection techniques, identifies which predictor variables are most important for explaining the variation in the response variable. This technique also allows for the identification of a best model and creates regression coefficients based on all the possible models (model-averaged coefficients), which are more robust than typical regression coefficients.

Assessing Model Fit. To determine which student variables best predicted student responses on the survey, we used the Akaike's information criterion corrected for small sample sizes (AICc) with regressions or mixed-effects models. The smaller the AICc values, the better the fit of the model to the data after a penalty for each additional variable present in the model. These AICc values are then used to rank the models. AICc differences (Δi), and Akaike weights (ω_i) are also calculated. These values represent two methods of assessing the strength of evidence supporting each model as the best model. The Δi encompasses the strength of evidence in support of each model as the best model: it represents the difference in AICc value between the best model and the current model being examined. The larger the Δi, the less likely the model. A common rule of thumb is that models with a $\Delta i > 10$ are considered poor predictors compared with the best model, and models with a $\Delta i < 2$ are considered statistically equivalent (Burnham and Anderson, 2003). A second way to evaluate support for a particular model is to compare Akaike weights (ω_i). These weights are a calculation of the likelihood of the observed data given a particular model that has been standardized so the sum of all the model weights adds up to one. These weights make it easier to compare models, as the likelihood is approximately the probability that the model is actually the best model given our set predictor variables. These weights can also be used to estimate the probability that a particular predictor variable is in the best model (relative variable importance). This relative variable importance is calculated by summing the Akaike weights across all the models that include a particular explanatory variable. Model-selection analyses were implemented in R using the MuMIn package (Barton, 2015). Even though one method may show that the models are statistically equivalent, the second method may indicate that one model is slightly better. In this case, there is not strong evidence to differentiate between these models.

Assessing Predictive Importance of Student Variables. In addition to identifying the best model, the multimodel inference approach also allows us to use information from all possible models to generate regression coefficients through model averaging (Anderson, 2008; Garamszegi, 2011). Similar to any regression coefficient, model-averaged regression coefficients inform us of the magnitude of the effect of each predictor variable and, combined with their standard error, can provide a measure of the likelihood that this effect is due to chance. The model-averaged coefficients differ from other regression coefficients in that they incorporate the underlying uncertainty (that is always present, although we often ignore it) as to which model best fits the data.

Model-selection creates multiple outputs, including tables of model rankings, model-averaged regression coefficient tables and statistics, and regression coefficients and statistics based on the best model. For the sake of brevity, the Results section of each study will present only two of these outputs: 1) written summaries of the model rankings and relative variable importance, which together indicate which explanatory variables most reliably predict the outcome variable across all models, and 2) written and visual summaries of the model-averaged regression coefficients. We chose to present the model-averaged results, because we did not have a clear best model for the majority of analyses, which is not uncommon for this type of observational study (Burnham and Anderson, 2003). In addition to the written summaries in the main body of the paper, we provide the ranked model and the model-averaged regression coefficients table in the Supplemental Materials.

Study I: What Roles Do Students Prefer to Play in Peer Discussions?

Methods. Students in terms A and B completed an open-ended question online after a class session with extensive peer discussions. This question was piloted in think-aloud interviews with undergraduates to determine whether students were interpreting the question correctly, and no alternative interpretations were observed. The question prompted them to describe their "preferred role in group work" and 313 students (88% of the class) provided a response.

A combination of grounded theory and content analysis was used to code student responses, which were stripped of identifying information (e.g., gender, ethnicity, name) before being coded (see Supplemental Materials section B for detailed methods). Through this method, preferences for four roles were identified: leader/explainer, collaborator, listener, and recorder (Supplemental Material Table 2). The leader/explainer category included any students whose responses indicated that they preferred to control the direction the group moved in and/or that they liked to explain answers to their groups with the caveat that they never indicated that

they listened to their groupmates' contributions. Collaborators were students who indicated they both preferred to listen and explain or asked questions during group discussions or were students who mentioned they preferred groups in which all students contributed equally. Listeners were students who said that they preferred having material explained to them but never mentioned they preferred to do any of the explaining. Recorders were students who exclusively said they liked to write down the group responses but did not say they listened or indicated they were a leader or explainer in the group. Student responses that that did not fit into one of these four categories and that were too uncommon to deserve their own category were binned into an "other" category.

Statistical Analyses. To identify the student characteristics that best predict the roles that students prefer to play in peer discussions, we used a model-selection approach paired with a multinomial regression. Only students with a complete set of predictor variables were included in this analysis ($n = 297$). In this analysis, we were limited in the number of predictors we could include in our model due to sample size constraints. Including an interaction between gender and race/ethnicity/nationality and our other predictor variables would have produced 202 potential models to test, which is almost as many models as we had data points. At the minimum, the number of models should not exceed the sample size for model selection (Burnham and Anderson, 2003). To be conservative and reduce the number of models, we chose to test only the main effects of our variables, which produced 32 potential models. This was substantially lower than our sample size. Thus, the full model we tested was: preferred role ~ gender + race/ethnicity/nationality + relative biology competency + friend in group + term.

Multinomial regression compares the odds of being in one category with the odds of being in each other category (Fox and Weisberg, 2010). For this analysis, once we identified the best model through model selection, we ran four multinomial regressions to compare all possible combinations of roles. The analysis was implemented in R using nnet (Venerables and Ripley, 2002) and MuMIn (Barton, 2015).

Results
Descriptive Statistics. The majority of the students in our analysis preferred to play a role in peer discussions that involved at least some explaining (Supplemental Material Table 2). The most common role that students preferred to play in peer discussions was the role of collaborator (both listening and explaining: 44.1%). The second most common role was leader/explainer (27.2%). The other two roles, which did not involve any explaining, were less common as preferences (exclusively a listener: 11.1%; exclusively a recorder: 5.0%).

Model Selection Results. The 95% confidence set of models contained 11 models, but the top three models contained the majority of the support (summed $\omega = 0.64$; Supplemental Material Table 3) and all had a $\Delta i < 2$ (indicating all three had equivalent support). Using Akaike weights, the first-ranked model ($\omega_i = 0.29$) was 1.32 times more likely to be the best model than the second model ($\omega_i = 0.22$) and was 2.2 times more likely to be the best model than the third model. Neither of these model-selection metrics provides strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

All three of the top ranked models contained gender. The first model contained two additional variables: friend in group, and race/ethnicity/nationality. The second best model differed from the first model only by the lack of race/ethnicity/nationality and the third best model lacked friend in group.

Relative Variable Importance Results. The variables gender and friend in group had the strongest evidence suggesting that they are in the best model (relative variable importance of 1 and 0.71, respectively). Although the race/ethnicity/nationality variable was present in the best and third best models, it had weaker support overall for being in the best model (relative variable importance = 0.51). Relative biology competency had the weakest support at 0.20.

Model-Averaged Multinomial Results. Of all the pairwise comparisons of preferred roles, the gender variable was only a significant predictor in comparisons with the role of leader/explainer (for full comparisons and p values see Supplemental Material Table 4). In all the pairwise comparisons, women were less likely than their male peers (with the same relative biology competency, status of friend in group and in the same class) to report preferring to play a leader/explainer role relative to any of the other roles. Women were four times more likely to prefer to be collaborators relative to leaders/explainers compared with their matched male peers ($p = 0.0001$). Although fewer students preferred to be listeners or recorders, women were 3.6 times and 11.2 times more likely to be in these categories than their matched male peers ($p = 0.005$ and 0.0025, respectively). Using predictions from the highest ranked model, the average male student in this study had a 45% chance of preferring to be a leader/explainer, whereas the average female student had only an 18.5% chance of preferring to be a leader/explainer (Figure 1A).

The race/ethnicity/nationality variable was primarily significant in comparisons involving the listener role (for full comparisons and p values see Supplemental Material Table 4). Asian-American students relative to their white American peers were more likely to prefer to be listeners than any other role (2.3 times less likely to prefer to be a leader/explainer, $p = 0.037$; 2.4 times less likely to prefer to be a collaborator, $p = 0.041$; 8.6 times less likely to prefer to be a recorder, $p = 0.009$). Underserved American and international students showed a similar pattern in comparison with white students: they were more likely to prefer to be listeners relative to most other roles, although not all p values reached the 0.05 threshold for significance (Supplemental Material Table 4). The only exception to this pattern was listener versus recorder for international students, since no international students ever preferred to be recorders. Using predictions from the highest ranked model, the average white student had a 3.9% chance of preferring to be a listener, the average Asian-American student had a 13.4% chance, the average underserved American student had a 21.6% chance, and the average international student had a 32.7% chance (Figure 1B).

Our control variable for relative biology competency seemed to be an important predictor for comparisons involving the listener role as well, although only one of these comparisons reached the $p < 0.05$ threshold. As biology competency increased, students were 1.6 times more likely to prefer to be leaders/explainers ($p = 0.027$), 1.4 times more likely to prefer to be collaborators ($p = 0.09$), and 2.0 times more likely to prefer to be recorders ($p = 0.051$) relative to listeners.
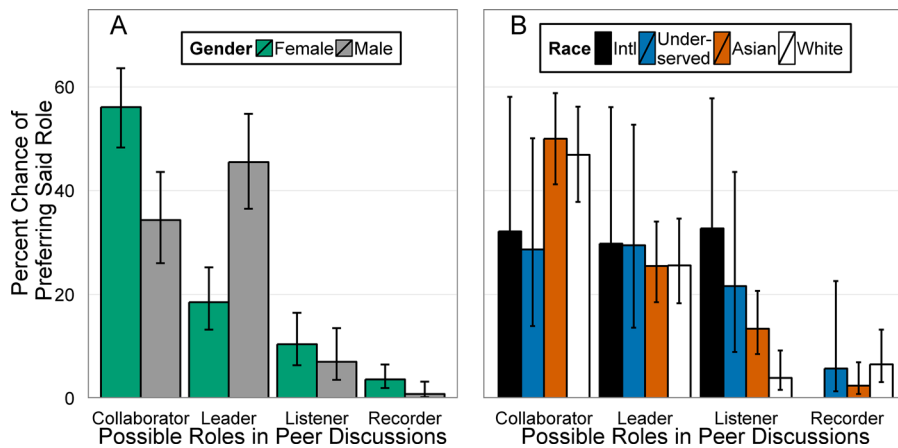
Finally, friend in group significantly predicted a student preference in only one comparison: collaborator versus listener. Students with a friend in the group were 2.5 times more likely than students without a friend in their group to prefer the role of collaborator to listener.

Conclusion Study I. In this study, we explored a student's preferred role in peer discussions. We identified four primary roles that students prefer to assume in peer discussions: leader/explainer, collaborator (listening and explainer), listener, and recorder. The majority of the students in our analysis preferred to play a role in peer discussions that involved at least some explaining, with the most common preferred role being collaborator. The second most common preferred role was leader/explainer.

Of the student identities we compared, gender had the most support for being in the best model. Gender predicted whether a student reported preferring to be a collaborator or a leader/explainer, with males more likely to prefer the leadership/explainer role and females more likely to prefer to be collaborators. Race/ethnicity and nationality influenced whether or not students preferred to do any explaining in their groups. Underserved American, Asian-American, and international students were all more likely to prefer playing listening roles compared with white students. This variable only had a 51% chance of being in the best model, however.

It is important to remember as we consider these results that these data indicate only the roles that students report preferring to play in their groups, which we propose may be a proxy for the actual role students play in peer discussions, on average, but we do not have data on actual participation. It is possible that students do not always get the opportunity to assume their preferred roles, so these results should be interpreted with some caution.

Study II: Do Students Indicate That They Are Excluded from Peer Discussions by Groupmates (Hypothesized Barrier 1)?
Methods
Open-Response "Worst Part of Group" Question. To explore how frequently students perceive other students as limiting their participation in group work, we asked students in terms A and B the open-ended question "What is the worst part about working in a group?" We intentionally chose a nondirect prompt to avoid priming students to talk about group dynamics; we conducted think-aloud interviews on this question to ensure that students were interpreting it the way we had intended. Students completed this question online the evening after a class session with extensive peer discussion. A combination of grounded theory and content analysis was used to code the 278 student responses (77.2% of enrolled students) received for the open-ended question. If the responses were too vague for coders to identify which category they belonged in (e.g., one word answers) or if students wrote something positive about peer discussions in response to this question, then those students were not included in the analysis (6.2% of responses).

Likert-Scale "Dominator in Group" Question. In addition, 323 students (89.7% of students in the class) in terms A and B responded to the following Likert-scale question prompt about dominating students: "There was one (or more) person in my group who dominated most of the discussion." The question had four response options ranging from strongly agree to strongly disagree. On all the questions, the most positive responses (strongly disagree there was a dominator) were scored as a 4 and the least positive as a 1 (strongly agree there was a dominator).

Statistical Analyses
Open-Response "Worst Part of Group" Question. We used logistic regression paired with model selection to determine which student-level variables best predict the likelihood of a student reporting a groupmate limiting their participation in peer discussion (1 = reported, 0 = did not report). The reference level for this logistic regression was "did not report" a groupmate limiting participation. This category included responses about anything other than groupmates limiting participation. Only students with a complete set of predictor variables were included in this analysis ($n = 278$).

The full logistic regression model used for model selection, which produced 52 possible models, was: external factors limit participation ~ gender + race/ethnicity/nationality + relative biology competency + friend in group + term + gender × friend in group + race/ethnicity/nationality × friend in group.

We were unable to include an interaction between gender and race/ethnicity/nationality, because including this variable led to perfect separation in our data analysis (Albert and Anderson, 1984). Perfect separation can occur when a data

set is small and the event being observed is rare. One option for dealing with perfect separation is to reduce the number of variables in the model. We found that taking out the gender and race/ethnicity/nationality interaction allowed the model to run.

Likert-Scale "Dominator in Group" Question. We used an ordinal regression to determine the odds of reporting that someone in the group dominated discussions. Our sample size for this analysis is small, which limits the number of predictors we can reliably include in our model. We chose to exclude the gender × race/ethnicity/nationality interaction variable, because a previously published study on this student population did not find this interaction to be significant (Eddy *et al.*, 2014). Only students with a complete set of these variables were included in this analysis ($n = 308$). The full logistic regression model (resulting in 72 possible models) was: dominator ~ gender + race/ethnicity/nationality + relative biology competency + friend in group + term + gender × friend in group + race/ethnicity/nationality × friend in group.

Results: Open Response to "Worst Part of Group Work" Question

Descriptive Statistics. We identified nine overall themes in the responses to the "worst part of group work" question. These included: lack of knowledge among students in group (24.1%); groupmate deficit (identifying traits of group members to explain their low participation, such as not knowledgeable, lazy, shy; 12.5%); groupmate(s) somehow preventing students from participation in discussion (12.5%); groupmates with conflicting ideas (9.9%); personal discomfort participating (9.9%); and in general, prefer lecture to peer discussions (6.8%). A multitude of other answers were provided, but these responses were too low in frequency to assign to individual categories, thus the last category is "other" (33.8%). A table of representative quotes for each category and a fuller description of these codes can be found in Supplemental Material Table 5.

Because we were specifically interested in identifying how groupmates might limit participation, we further disaggregated student responses for this category. We identified three distinct means by which students felt groupmates limited their participation in peer discussions: 1) one student dominated group discussions (54.4% of students who reported groupmates prevented participation listed this as the worst part of group work), 2) fear of groupmate reactions to one's contribution (25.6%), and 3) group members moved through discussion too quickly for other members (20%).

Overall, the number of students who reported the barrier of groupmates limiting their participation was low. Thus, for statistical power, we reaggregated the three ways students were excluded by groupmates into one general category of "groupmates limited participation." All the other codes were combined into the category "did not report a groupmate limiting participation." Thus, a binary outcome variable identifying whether or not a student reported a groupmate limiting participation was created from the open responses.

Model-Selection Results. The 95% confidence set of models contained 27 models. The top five models had the majority of the support (summed ω = 0.54; Supplemental Material Table 6). Of these models, four had equivalent support

($\Delta i < 2$) for being the best model. Using Akaike weights, the first-ranked model ($\omega_i = 0.17$) was almost equally as likely as the second-ranked model ($\omega_i = 0.14$). Together, these two models were 3.9 times more likely to be the best model than the third-ranked model and 5.17 times more likely than the fourth-ranked model. Neither of these model-selection metrics provide strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

The top two models both contained relative biology competency. The second model also included gender.

Relative Variable Importance Results. No variable in the full model had a 100% chance of being in the best model. The highest support was for the control variable relative biology competency (relative variable importance = 0.76), which was in all four of the equivalent models ($\Delta i < 2$) and in 59% of the 95% confidence set of models. Gender had the next highest support for being in the best model at 48%. Both friend in group and term had some support for being in the best model (relative variable importance = 0.30 and 0.19, respectively). The interactions between friend in group and gender, as well as friend in group and race/ethnicity/nationality, had very little support for being in the best model (relative variable importance = 0.05 and 0.01, respectively).

Model-Averaged Logistic Regression Results. The overall rate of reporting that groupmates limited participation was low; however, we found that relative biology competency impacted this rate relative to other concerns ($\beta = -0.38 \pm 0.182$; $p = 0.037$; Supplemental Material Table 8). All else being equal, students who were one SD below the mean biology competency were 1.5 times more likely to report that a groupmate limited participation than students at the mean biology competency. Students who were two SDs below the mean were twice as likely to report this. Although several other variables had regression coefficients comparable in magnitude or larger, the SEs associated with those variables were large enough that none of the other variables included in the full model were significant at the $p = 0.05$ level, nor were there any trending patterns ($p \leq 0.1$; Supplemental Material Table 8).

Results: Likert-Scale "Dominator" Question

Descriptive Statistics. When specifically prompted with a question about dominators, 29.4% of students agreed or strongly agreed that someone in their group dominated the discussion.

Model-Selection Results. Twelve models were in the 95% confidence set of models. The top two models contained the majority of the support (summed ω = 0.60; Supplemental Material Table 7), and the top three models had fairly equivalent support for being the best model ($\Delta i < 2$). Using Akaike weights, the first-ranked model was 1.9 times more likely to be the best model than the second model and 2.4 times more likely than the third model. Neither of these model-selection metrics provide strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

All three equivalent models contained relative biology competency, race/ethnicity/nationality, and term. The first model contained only these variables. In addition to these variables, the second model contained friend in group, and the third model contained gender.

Relative Variable Importance Results. Relative biology competency, race/ethnicity/nationality, and term all had strong

support for being in the best model (relative variable importance > 0.96; Supplemental Material Table 8). Friend in group and gender had moderate support (relative variable importance 0.39 and 0.32, respectively). The three interaction terms had little evidence supporting their inclusion in the best model (relative variable importance was less than 0.05 for each).

Model-Averaged Ordinal Regression Results. Across all our possible explanatory variables, only relative biology competency, race/ethnicity/nationality, and term significantly predicted a student's response to the question of whether there was a dominator in the group (Supplemental Material Table 8).

Student competency with the course material predicted whether students reported a dominator in their group. As a student's Relative Biology Competency increased, he or she was 1.7 times more likely to report no dominator in the group ($\beta = 0.51 \pm 0.125$; $p = < 0.0001$). Thus, the best model predicted that the average student whose relative biology competency placed him or her one SD above the mean had a 28% chance of reporting strongly agreeing or agreeing that there was a dominator, whereas students one SD below the mean had a predicted 38% chance of reporting this, and students two SDs below the means had a 51% chance.

Race/ethnicity/nationality also predicted student responses on this question. International students were 4.8 times more likely to report a dominator in their group than white American students ($\beta = -1.48 \pm 0.499$; $p = 0.00005$). In addition, there was a strong trend for Asian-American students to also be more likely to report a dominator ($\beta = -0.52 \pm 0.288$; $p = 0.069$) relative to white American students. The best model predicts that a student, average in all other ways, who was international or Asian American had a 52 and 30% chance, respectively, of strongly agreeing or agreeing that there was a dominator in his or her group (Figure 2). The average student who was a white American had a 21% chance.

Conclusions Study II

In Study II, we explored whether students reported that groupmates limited their participation in peer discussions.
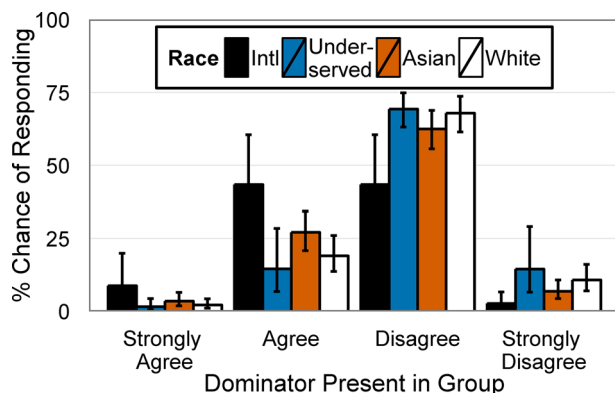

Figure 2. Race/ethnicity/nationality predict which students report a dominator in their peer discussion group. Model prediction based on the ordinal regression model with the most support: dominator present (Y/N) = relative biology competency + race/ethnicity/nationality + term. Models were considered statistically equivalent, so we chose the model with the slightly higher Akaike weights for this figure.

Overall, we found that only a small number of students reported this barrier occurring and that a student's gender did not predict how a student would respond to questions about this barrier. In addition, a student's race/ethnicity/nationality was only a significant predictor of student responses when students were specifically prompted to think about dominators. Specifically, international students and Asian-American students were both more likely than white American and underserved American students to report a dominator.

Study III: Do Students Express Anxiety with Participation (Hypothesized Barrier 2)?

Methods. In this study, we compared students' anxiety related to participation in small peer discussion groups and in front of the whole class. To document this anxiety, we used a previously published survey (Micari and Drane, 2011) that was developed to explore intimidation in peer discussions in college-level STEM classrooms. The original survey had 10 Likert-scale questions about intimidation that the authors had developed based on conversations with students and faculty. These questions sorted onto two factors: comfort being oneself (a measure of whether students felt they could be imperfect in their peer discussion group without negative effects) and social comparison concern (concern about one's abilities relative to one's groupmates' abilities). In this previous study, social comparison concern was found to be a predictor of both course grade and self-reported participation in peer discussions.

We used the 10-question survey to assess students' experiences in peer discussions. To address the student experience of talking in front of the whole class, we modified the same 10 questions by removing references to "peer discussions" and replacing them with "whole-class discussions" (see Supplemental Material Table 9). Think-aloud interviews were conducted with students before the survey was administered to ensure that students were interpreting the questions appropriately. Students in term C were given the 20-question Likert-scale survey during week 7, and 88.6% ($n = 624$) of the class completed the entire survey. Each item on the survey had four options ranging from "strongly agree" to "strongly disagree." On all the questions, the most positive responses (lowest social comparison concern or greatest comfort) were scored as 4 and the least positive as 1. Therefore, a large number on either factor indicates a more positive experience, and a lower number indicates a more negative experience.

We ran initial tests on both the reliability of these two factors and on differential item functioning. Cronbach's alpha values for the two survey factors for both peer discussions and whole-class discussion were greater than 0.71. Four questions on the 20 questions demonstrated differential functioning by gender. Two of these favored females and two favored males. Results for these preliminary analyses can be found in Supplemental Material section C1.

Statistical Analyses. Our survey required students to answer the same 10 questions in two different contexts: one in reference to peer discussions and one in reference to whole-class discussions. Thus, we had repeated measures on each student. To account for these repeated measures, we use general linear mixed-effects models with a random effect variable (1|Stu.ID) that allows clustering of responses by student

(Zuur, 2009). Linear mixed-effects models were analyzed in R using the lme4 package (Bates *et al.*, 2014).

To identify the subset of student-level variables that predicted a student's response to these two factors (comfort being oneself and social comparison concern) in peer and whole-class discussions, we used model selection paired with a linear mixed-effects model. The variables used in this analysis were different from those used in Studies I and II. We did not include the variable friend in group, because that was not relevant to the whole-class context. We added a variable for participation context (whole-class vs. peer discussions) and an interaction between these contexts and gender (gender × participation) and race/ethnicity/nationality (race/ethnicity/nationality × participation). Only students with a complete set of these variables were included in this analysis ($n = 623$). Thus, our initial full model (producing 36 possible models) used for model selection for each factor was: response = relative biology competency + gender + race/ethnicity/nationality + participation context + gender × race/ethnicity/nationality + gender × participation context + race/ethnicity/nationality × participation context + (1|Stu.ID).

Results Factor 1: Comfort Being Oneself

Descriptive Statistics. Across the four questions in the comfort factor for peer discussion, the median response of students was a 12 ± 2.1 SD (out of a possible 16), an average response of agree on the question. In whole-class discussion, the median response of students was 10 ± 2.4, which indicates a neutral response to comfort (between agree and disagree).

Model-Selection Results. The 95% confidence set of models contained four models (Supplemental Material Table 10). Using Akaike weights, the top two models contained the majority of the support ($\omega_i = 0.67$), although the top three were all considered equivalent ($\Delta i < 2$). The first model was only 1.2 times more likely than the second-best model. Neither of these model-selection metrics provide strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

The first model contained gender, relative biology competency, participation context, and gender × participation context. The second-best model contained all of the variables of the best model and, additionally, race/ethnicity/nationality variables and interactions. All three of the best models contained gender, relative biology competency, participation context, and the gender × participation context interaction.

Relative Variable Importance Results. The variables relative biology competency, gender, participation context, and gender × participation context had the highest relative variable importances (1.00) across the possible models (Supplemental Material Table 12). Although not present in the best model, race/ethnicity/nationality and race/ethnicity/nationality x participation also had moderately strong support for being in the best model (0.64 and 0.52, respectively).

Model-Averaged Linear Mixed-Effects Model Results. Relative biology competency and participation context were the only main effects that significantly predicted student responses on the comfort being oneself factor (Supplemental Material Table 12). In addition, three interaction terms were also predictive. Across both participation contexts, students reported a 0.51 ± 0.068 increase in comfort being themselves for each SD increase in relative biology competency ($p < 0.0001$). Thus, students who were two SDs below the mean biology competency reported a level of comfort being oneself that was half of an SD lower than the average comfort in the class.

Participation context also impacted student responses on the comfort being oneself scale, but this impact was moderated by gender and race/ethnicity/nationality. Overall, participation in front of the whole class caused male students to report a 2.3 point decrease in comfort being oneself relative to small-group work ($p < 0.001$; Figure 3A); this reduction was more than an SD on the comfort being oneself factor. For women, it was a 3.41 reduction ($\beta = -1.1 \pm 0.218$; $p < 0.0001$; Figure 3A). In addition, race/ethnicity/nationality was not in the best model, but there is some evidence in the model-averaged regression coefficient that suggests this variable could be predicting student responses, particularly for international students relative to white American students. The race/ethnicity/nationality variable was likely not included in the best model, because international students are such a small proportion of our sample (68 students or 5.8% of the students in the study). Overall, international students were more positive about their experiences in whole-class discussions, reporting 1.44 ± 0.485 ($p = 0.003$) more positively, which is more than half an SD increase in the response.

Finally, although gender did not have a main effect on student comfort (-0.003 ± 0.254; $p = 0.99$), there was a significant interaction term between Asian-American and female status. In both contexts, Asian-American female students were more likely to report a lower level of comfort than Asian-American male students (-0.70 ± 0.293; $p = 0.017$), whereas there was no difference between how students of different genders reported for the other groups. Again, this variable had only moderate support for being in the best model.

Results Factor 2: Social Comparison Concern

Descriptive Statistics. Across the six questions related to social comparison concern, the median response of students was 17 ± 3.0 SD, indicating a slightly positive average response in peer discussions. In whole-class discussions, the average
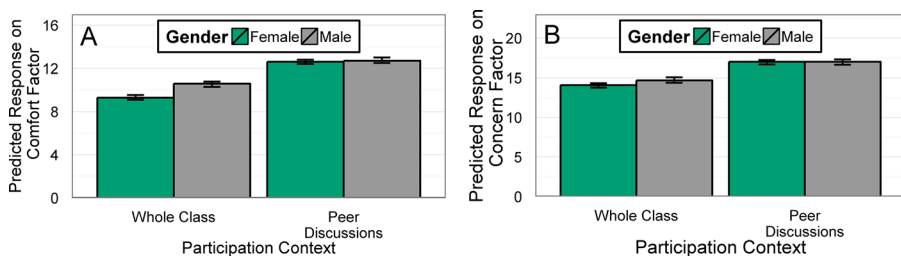

Figure 3. Gender × participation context predicts student responses on (A) comfort being oneself and (B) social comparison concern factors. Predictions based on best model for the factors: response = gender + relative biology competency + participation context + gender × participation context + (1|Stu.ID). Models were considered statistically equivalent, so we chose the model with the slightly higher Akaike weights for these figures.

response was 15 ± 3.1 (out of a possible 24) on the questions on social comparison concern in large groups, indicating an overall neutral response to social comparison concern questions.

Model-Selection Results. Five models made up the 95% confidence set of models (Supplemental Material Table 11). Two of these models had equivalent support ($\Delta i < 2$) for being the best model as well as the majority of support ($\omega_i = 0.74$), although, using Akaike weights, the first-ranked model was twice as likely to be the best model as the second-ranked model. Neither of these model-selection metrics provide strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

Both models contained relative biology competency, participation context, gender, and gender × participation context. The best model also includes the race/ethnicity/nationality variable and its interactions with participation context and gender.

Relative Variable Importance Results. Four variables had very strong evidence for being in the best model (> 0.93): relative biology competency, participation context, gender, and gender × participation context (Supplemental Material Table 12). Race/ethnicity/nationality and race/ethnicity/nationality × participation context, and race/ethnicity/nationality × gender all had moderate support for being in the best model (0.74, 0.68, and 0.57, respectively).

Linear Mixed-Effects Model Results with Model-Averaged Regression Coefficients. Relative biology competency, participation context, and race/ethnicity/nationality were all main effects that significantly predicted student responses on the social comparison concern factor (Supplemental Material Table 12). As students' relative biology competency increased, they reported lower social comparison concern ($\beta = 1.27 \pm 0.009$; $p < 0.0001$). Thus, students with a relative biology competency 1 SD below the mean are predicted to report social comparison concern that is two-fifths a SD greater than students with the mean relative biology competency.

Participation context also significantly predicted student social comparison concern responses, but it was moderated by several other factors (Figure 3B). Overall, male students reported lower levels of social comparison concern in peer discussions relative to whole-class discussions ($\beta = -2.44 \pm 0.270$; $p < 0.0001$). With the significant interaction between gender and participation context, female students reported slightly higher levels of social comparison concern than males in whole-class discussion ($\beta = -0.68 \pm 0.225$; $p = 0.002$). In addition, race/ethnicity/nationality mediated the relationship between participation context and student response. International students reported higher levels of social comparison concern in peer discussions relative to white American students ($\beta = -1.84 \pm 0.751$; $p = 0.014$). In contrast, a significant interaction between international status and participation context led to international students responding similarly to white American students on this factor for whole-class discussions ($\beta = 1.44 \pm 0.485$; $p = 0.001$).

Conclusion Study III. In this study we explored whether students reported anxiety with participation in peer and whole-class discussions. Overall, students seem to be less anxious in small peer-group discussions as evidenced by the responses for both the comfort being oneself and social comparison concern factors. This pattern is true for all white, Asian, and underserved American students, but particularly true for females. Females feel an equally low level of anxiety as their male peers in peer discussions but disproportionately more anxiety than males in whole-class discussions. Finally, although there was less support for including race/ethnicity/nationality in the best models, the model-averaged regression coefficients suggest that international students may be less comfortable in peer discussions relative to their comfort in whole-class discussions compared with their white American peers.

Study 4: Do Students Report Not Valuing Peer Discussions (Hypothesized Barrier 3)?

Methods. In this study we explored how aspects of the group experience, student social identities, and relative biology competency influence the value that students perceive in peer discussions. We used seven questions from the Micari and Drane (2011) survey used in Study III that focused on comfort and confidence participating in peer discussions. We also developed 10 additional questions. Five of these questions explored how well a student perceived his or her group to function. The other five explored a student's perception of how different aspects of peer discussions influences his or her understanding of a course topic. These five questions were conceptualized by the authors as "value of peer discussions" for learning (for full list of questions see Supplemental Material, Supplemental Analyses for study 4). The new questions were developed based on student comments and concerns expressed in focus groups conducted in a previous term of this course (unpublished data). Think-aloud interviews were done with students to check whether students were interpreting the questions correctly. We found that students were able to explain their answers in ways consistent with the intent of the questions. A longer survey capturing student classroom engagement, which includes these questions, is currently being developed (unpublished data). Factor analysis was implemented in Mplus.

Students in terms A and B took this survey online after peer discussions in class in week 7 of the course.

Statistical Analysis

Exploratory Factor Analysis. We used principal components analysis to determine the number of factors this survey represented. In addition to the response factor related to the value of group work, the principal components analysis, using Varimax with Kaiser normalization rotation method, produced two factors that accounted for 75.61% of the overall variance and were conceptually appropriate. Factor 1 was conceptualized by authors as "group function," and included five questions focused on groupmate contributions and interactions (see Supplemental Material Table 13 for questions and factor loadings). The median response on factor 1 was 15.0 (out of a possible 20), indicating on average a somewhat positive attitude about how well a student's group was functioning. Factor 2 included the seven questions from the Micari and Drane (2011) survey and was conceptualized by the authors as "comfort and confidence with contributions to group." The median response on factor 2 was 20 (out of a possible 28), indicating an average positive response to this factor. For analyses of how demographic variables impact Factors 1 and 2 see Supplemental Materials Section E.

Model Selection. In addition to our student variables, we included the factors of group function and comfort and confidence as explanatory variables for predicting student responses on the value of peer discussion scale. This increased the potential variables in our model to 10, which would have produced 288 models, which was more than our sample size ($n = 276$). Due to the weak support for gender × race/ethnicity/nationality and race/ethnicity/nationality × friend in group in our previous models, we chose to exclude these variables from our analysis. This reduced the number of possible models to 160. Thus, the full model was: value of group work = relative biology competency + gender + race/ethnicity/nationality + friend in group + group function + comfort and confidence + term + gender × friend in group.

In a preliminary analysis, we explored the possibility that the comfort and confidence, group function, and relative biology competency could be correlated, because we saw a relationship between biology competency and these factors in Study III. We found that relative biology competency had a small correlation with both group function and comfort and confidence (~.25), but group function and comfort and confidence were fairly strongly correlated with each other ($r = 0.64$). With this concern, we tested whether collinearity would be a problem in the full model and found no evidence of a problem (variance inflation factors all below 2).

Results

Descriptive Statistics. Across the five questions of the value of peer discussion factor, the median student response was a 14 ± 2.33 SD (out of a possible 20). This indicates a fairly positive response by students overall (an average response just below "agree" across these questions).

Model-Selection Results. Seventeen models were included in the 95% confidence set of models (Supplemental Material Table 17). Two of these models had essentially equivalent support for being the best model ($\Delta i < 2$). Of these two models, the first-ranked model was only 1.5 times more likely to be the best model relative to the second-ranked model. Neither of these model-selection metrics provides strong support for a single model, so we focus on the model-averaged results that account for the uncertainty about which model is the best model.

The two models differed only in whether or not they contained the comfort and confidence with participation factor. Otherwise, they both contained the group function factor, friend in group, relative biology competency, term, gender, and friend in group × gender. Overall, the top model explained 29.7% of the variation in student responses on the value of peer discussions factor.

Relative Variable Importance Results. Three variables had very strong support for being in the best model: group function, relative biology competency, and friend in group

(all > 0.98; Supplemental Material Table 18). Fairly strong support for being in the best model was found for term, gender, gender × friend in group, and comfort and confidence with participation (0.88–0.59). Finally, there was only slight support for race/ethnicity/nationality being in the best model.

Linear Regression Model-Averaged Regression Coefficients. Three main effects significantly predicted a student's response on the value of peer discussions factor: response on group function factor, relative biology competency, and term (Supplemental Material Table 18). For every unit increase on the group function factor, students reported 0.51 ± 0.089 more positively on the value of peer discussions scale ($p \leq 0.0001$, Figure 4A). Thus, a five-unit increase on the group function scale would increase the value of peer discussion by a little more than 1 SD. Relative biology competency demonstrated the inverse relationship: as competency increased, students saw less value in group work (-0.50 ± 0.140; $p = 0.0004$). In addition, there seems to be some term by term variation in the value of peer discussions (0.71 ± 0.293; $p = 0.015$).

In addition to these main effects, there was a significant interaction between gender and friend in group. Female students without a friend reported perceiving less value in peer discussions than their male peers (-1.05 ± 0.579; $p = 0.068$; Figure 4B). However, this pattern disappeared when female students were working with at least one friend: females reported seeing equal value in peer discussions as their male peers (1.42 ± 0.248; p = 0.014).

Conclusion Study 4. Both students' experiences in their peer groups and their characteristics predicted their response on the value of peer discussions factor. Specifically, students' perceptions of how well their groups function and whether or not a friend was present in their group predicts their response for value of peer discussions. As group function increased, students perceived more value in peer discussion. The impact of having a friend in the group was specifically predictive for female students: females without a friend reported lower value in peer discussions relative to their male peers, but females reported slightly higher value in peer discussions when they had a friend in the group. The presence of a friend did not impact males.

DISCUSSION

The types and structure of class activities that instructors implement have an impact on the learning experience of students, which may help explain why not all active-learning classrooms show the same level of learning gains (Andrews *et al.*, 2011; Freeman *et al.*, 2014). In this paper, we focused on the experience of students in peer discussions, one of the
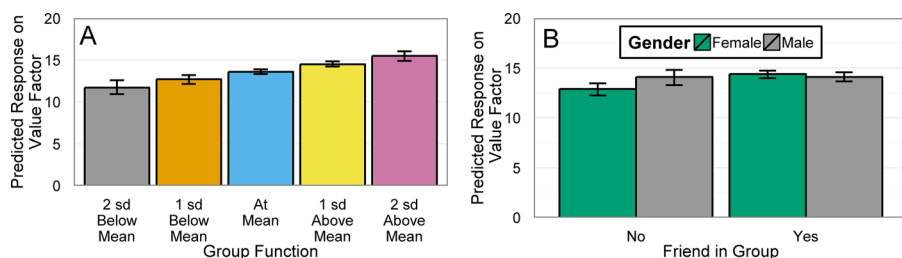

Figure 4. The impact of group function (A) and friend in group (B) on the perceived value of peer discussions. Model prediction from best model for the average student: value = function + comfort + friend + gender + term + relative biology competency + friend × gender. Models were considered statistically equivalent so we chose the model with the slightly higher Akaike weights for these figures.

predominant forms of active learning in large STEM classrooms. For students to benefit equally from peer discussions, it is likely that all students need to engage as both explainers and listeners during an activity so everyone in the group has an opportunity to engage constructively or interactively, leading to deeper understanding (Chi and Wylie, 2014). However, we found that students' self-reported preferred roles in peer discussions vary based on race/ethnicity, nationality, and gender. In addition, we found support for all three hypothesized barriers to participation in peer discussions (Table 1). Identifying potential barriers provides guidance as to how to more purposely structure peer discussions to increase equity in participation. This study is a first exploratory step in deepening our understanding of inequities associated with peer discussion, and further work is necessary to more fully define the link between these barriers and actual participation differences.

The Preferred Roles of Students of Different Genders and Races/Ethnicities/Nationalities and the Barriers That Might Underlie These Differences

In this exploratory study we identified that different student groups preferred to assume different roles during peer discussions (a possible proxy for the actual role assumed, on average, in peer discussions). We then identified the intensity of three hypothesized barriers to peer discussions to determine whether they could help explain these patterns of participation. The three hypothesized barriers were 1) students are excluded from peer discussion by groupmates; 2) students' anxiety about participation limits their participation; and 3) low student perceptions of the value of peer discussions limits their participation. Below we summarize whether each hypothesized barrier could possibly explain the differences in preferred roles reported in peer discussions by students of different genders or races/ethnicities/nationalities.

Gender. Female students reported preferring the collaborator role most frequently in small-group work (55% chance of reporting this role), whereas males preferred leading (45% chance of responding with this role) after controlling for race/ethnicity/nationality, biology competency, and other course-level variables. Females had an 18% chance of responding with the role of leader. These differences in role preference could be influenced by a multitude of factors beyond the scope of this paper, but of the three barriers we explored, only one demonstrated a similar variation between genders. Males and females had an equally low likelihood of reporting a dominator in their group, thus groupmates limiting participation (Barrier 1) does not seem to explain the differences in participation reported. In addition, both female and male students reported equally low anxiety concerning participation in small groups, so differences in anxiety about participation (Barrier 2) were not supported as a root cause of differential participation between the genders. The only observed difference between males and females was a difference in the perceived value of peer discussions for learning (Barrier 3): females, unless they were working with a friend, reported seeing lower value in peer discussion than their male peers. Further investigation of Barrier 3 may contribute to our understanding of the differences in roles that males and females prefer to assume in peer discussions.

Our finding of equal levels of anxiety in peer discussions for males and females differs from a prior publication in college-level STEM education. Micari and Drane (2011), the authors who developed the survey on comfort and social comparison concern used in our research, found that females were more apprehensive than males in peer discussions. However, their group sizes were larger (five to seven students versus two to four in our study), and the group work was part of a supplemental instruction program completed outside class, whereas the peer discussions that we studied were part of normal classroom practice. In addition, we studied only biology classrooms, whereas they studied STEM classrooms more broadly; only a small percentage (12%) of their students were enrolled in a biology course. Gender may be more salient in peer discussions in other STEM fields (Eddy and Brownell, 2016), where females are a numerical minority, which can cause increased anxiety for the underrepresented gender (Thompson and Sekaquaptewa, 2002).

Table 1. Summary of patterns in student responses for gender, race/ethnicity, and nationality relative to appropriate reference group

Student groups:	Participation patterns (measured by preferred role in peer discussions)	Hypothesized barriers to participation in peer discussions:		
		Students are excluded from peer discussion by groupmates	Students' anxiety about participation limits their participation	Student perceptions of the value of peer discussions limits their participation
Gender: reference level male	*Females*: decreased preference to be a leader/explainer relative to any other role	*Females:* no difference	*Females:* no difference (increased anxiety in whole-class discussions)	*Females without friend in group*: decreased value *Females with friend*: increased value
Race/ethnicity: reference level: white American	*Asian American:* increased preference to be listener *Underserved American:* increased preference to be listener	*Asian American*: more likely to report a dominator *Underserved American*: no difference	*Asian American*: No difference *Underserved American*: No difference	*Asian American*: no difference *Underserved American*: no difference
Nationality: reference level: white American	*International*: increased preference to be listener	*International*: more likely to report a dominator	*International*: increased anxiety in peer discussions (decreased or equal in whole-class discussions)	*International*: no difference

Race/Ethnicity. Underserved American and Asian-American students were more likely than white American students to prefer the role of listener, particularly relative to the leader/explainer role. Of the three barriers tested, none differentiated between white American and underserved American students, so we have no evidence to help explain the differences between these two groups. However, we did find evidence that Barrier 1 may contribute to the preferred role difference between Asian-American and white American students; Asian-American students were more likely to report that someone in their group dominated discussions. The results for underserved American students should be considered tentative, because our sample size is small given the composition of students at our institution, and our findings should be tested further at institutions with larger populations of underserved American students.

Nationality. International students were more likely than white American students to prefer the role of listener. Our results reveal that Barrier 1 may help explain this difference in preferred roles: international students were more likely than white students to report that someone in their group dominated discussions. There was also some evidence that Barrier 2 might explain some of the role differences: international students reported higher anxiety than white American students in peer discussions for comfort being oneself, one of the two measures of anxiety.

Relative Biology Competency. Although not necessarily a stable construct (i.e., students' feelings of competency and actual competency in the course may vary across the term), it is interesting to consider the impact of our control of relative biology competency on students' responses. Student performance on exams relative to their peers did not impact the role that they preferred in peer discussions. However, we do find evidence that competency may impact the role that students may actually play in peer discussions, as students with low relative competency were more likely to report that 1) their groupmates limited their own participation in some way, 2) there was a dominator in their group (Barrier 1), and 3) they had more anxiety about participation (Barrier 2). Surprisingly, despite experiencing these barriers to a greater degree than students with higher biology competency, students with lower than average biology competency reported seeing more value in peer discussions than students with higher relative competency. Thus Barrier 3 (value) does not seem to explain why students with different levels of biology competency prefer to play different roles in peer discussion. These Barrier 3 results are perplexing, because, as listeners, these students with lower competency are likely not gaining the maximal learning benefit of engaging in the discourse that peer discussion is designed to offer (Chi and Wylie, 2014). In contrast, high-ability students who have been shown to gain the most from peer discussions (Beichner *et al.*, 2007; Jensen and Lawson, 2011) seem to value these discussions the least. It would be of interest for future studies to further investigate some of these suggestive patterns.

Summary. In sum, it seems that all three of our hypothesized barriers exist in the classroom during peer discussions, but not all factors were useful for explaining the preferred role differences between students of different social identities. Our study design constrains us to rely on comparing the alignment of preferred roles with patterns of self-reported experiences with barriers, which limits the scope of our conclusions. We cannot say that these barriers actually explain the differences in participation observed between students of different social identities and biology competencies, but we do provide a first exploratory picture of some of the factors influencing student experiences in peer discussion. Further work measuring the intensity and frequency of these barriers on a population of students with whom actual participation data are collected will be necessary to elucidate the possible mediating role of these three barriers.

Improving Equity in Active-Learning Classrooms

Before discussing recommendations, we need to acknowledge that we did not measure individual student learning or a product resulting from peer discussion. It is possible that student experience in peer discussions does not correlate with individual student learning or group productivity; students may learn in a group even when they are not comfortable, and there is some evidence that a group can be more productive when there is less equity (Anderson and Brown, 2010). There is a need for further studies to determine the relationship between equity in peer discussions in college STEM classrooms and student learning from these experiences.

Recommendation 1: Use Peer Discussions to Promote Gender Equity in Participation

Our study suggests that the use of peer discussions in class may increase equity in participation in large active-learning biology courses. We found that males and females experience similarly low levels of anxiety with participation in peer discussions (Study III). Although we cannot directly make the connection, we predict that increased comfort in peer discussions could indicate elevated willingness to participate in peer discussions relative to discussions in front of the whole class. Support for this causal relationship between anxiety and actual participation comes from observations of whole-class discussions. In whole-class discussions, female students are less likely to volunteer or be called on than their male counterparts, even though they make up more than half the class (Eddy *et al.*, 2014). This corroborates students' reported anxiety with participation in front of the whole class: females are disproportionately more anxious in this context. If anxiety predicts actual participation, as it seems to in whole-class discussions, then our results suggest that peer discussions should promote more equitable participation for males and females than whole-class discussions.

Recommendation 2: Consider Structuring Interactions in Peer Discussions to Improve Participation Equity

In our experience, most instructors using peer discussion in large STEM classrooms do not deliberately structure group interactions. Students typically self-select into groups and often are not provided explicit instructions beyond the initial question posed. Our results indicate unstructured group dynamics could generate inequities in the roles that students assume in group discussions.

A common explanation for why there are differences in participation in peer discussion follows a student deficit model: "lazy or unmotivated students" do not pull their weight in the group, such that the high-achieving students do all the work. However, our data raise the possibility that perhaps instead of students being lazy or unmotivated, students face barriers such as anxiety about group work, low perceived value of peer discussion for their learning, or contending with other students in the group who are dominating. Reframing inequities in participation in this way puts the onus on the instructor to structure the interactions in peer discussions to promote equal opportunities for all students to participate in the learning activity.

Many methods of structuring group interactions have been proposed in the literature (reviewed in Csernica *et al.*, 2002; Tanner *et al.*, 2003). Our data prompt us to suggest two important aspects for instructors to consider. First, some students seem to see more value in group work when they are able to work with people with whom they are friends. The common assumption that instructor-assigned groups of random students are better (Jacobs and Hall, 2002; Roseth *et al.*, 2008) is in direct contradiction with our data and may preferentially benefit males, exacerbating gender inequities. Furthermore, the practice of changing groups throughout the semester may hinder students who benefit more from sustained connections with others. In addition to letting students work with their friends, we need to recognize that not all students have pre-existing relationships with peers in their class. Instructors could promote the development of these relationships by assigning students to groups and helping students establish relationships with their groupmates by 1) keeping these groups for the duration of the term and 2) purposively incorporating time and activities to encourage students to build connections with groupmates. Even basic connections between students like discovering a shared a birthday can increase the performance of students in groups (Walton *et al.*, 2012).

Second, our data suggest that it is important to structure group interactions to prevent an individual member from dominating the peer discussions. This could be done through the structuring of the activity itself or through instructor actions and examples (Tanner, 2013). For instance, an instructor could require groups to rotate who talks first in each activity or who assumes which role in the group. A variation of this would be to have students work in dyads that enforce equal participation; one partner speaks for a period of time without the other student responding at all, and then they switch roles (Wickett, 2000). In addition, instructors could explicitly provide students with example questions or prompts that facilitate group discussions (Menekse *et al.*, 2013). Finally, assigning unique roles to each group member (Johnson *et al.*, 1998; Cohen *et al.*, 1999) or asking students to reflect on group dynamics (Kotsopoulos, 2010) may increase student awareness of how they are participating in a group.

The structure of the specific activities themselves can also influence equity in participation. Complex open-ended tasks involving uncertainty increase the need for a range of perspectives and problem-solving strategies, which encourages students to listen to one another and not just accept the first idea offered (Cohen *et al.*, 1999). In contrast, tasks with straightforward simple answers may not require the contribution of multiple students (Cohen *et al.*, 1999). Structuring

activities to promote positive interdependence may also enhance equity. Positive interdependence involves sharing a common goal and recognizing that success requires the participation of all individuals. Several activity formats promote positive interdependence; for example, jigsaw-style activities require each student to provide a unique piece of information to a task, which creates a situation in which the group needs everyone's participation to successfully complete a task (Tanner *et al.*, 2003). In contrast, some activity structures may lend themselves to promoting inequities. For example, requiring large groups of students to share one worksheet may limit who has access to the information contained on that worksheet. Similarly, having large groups of students share one set of materials to build a model may exacerbate inequalities between those handling the materials and those passively watching.

Limitations

In this study, we relied solely on student self-reported data from three iterations of a single course that were all taught using similar active-learning methods. This limits the scope of the conclusions we can draw. First, we do not know how accurately students are able to report their experiences in peer discussions. While it would be useful to pair direct observations of student groups with student self-reported values, the logistical challenges prevented us from doing so in these classrooms. Second, as stated previously, we do not know how well students' perceptions of their experiences align with actual learning gains or the quality of a group's product. Third, there is a spectrum of active-learning approaches, and the extent to which the instructor makes the classroom student-centered will likely impact the extent to which students feel comfortable participating in peer discussions and in front of the whole class. We anticipate that our results could change given a classroom with different types or intensities of active learning, including differences in group size (e.g., two versus six students per group). We encourage others to explore how the classroom environment established by the instructor impacts student participation. Fourth, a student's prior experience in active-learning or traditional lecture courses may influence his or her willingness to participate. It is possible that students in this study who had previously experienced active learning may see more value in it, but this also may be dependent on how effective that prior active learning was. Finally, this research was done at an R1 institution with a student body composed of primarily white and Asian-American students. We do not know whether the patterns we observed in this study would be similar in other institutional settings; therefore, it is important that similar research be conducted at different types of institutions.

CONCLUSIONS

We conclude that student race/ethnicity, nationality, gender, and relative biology competency can predict students' experiences with peer discussion. We provide evidence that three possible barriers to participation in peer discussions exist in biology classrooms: groupmates limiting participation, student anxiety with participation, and the value a student perceives in peer discussions. As instructors and education

researchers continue to evaluate the effectiveness of particular active-learning strategies, our data suggest it is important to disaggregate the data based on social identities and identify factors that may underlie these differences in classroom experience. How instructors choose to structure—or not to structure—their students' peer discussions may have a differential impact on participation, and only by identifying the barriers underlying participation differences can instructors design and implement learning interventions that are effective for all students.

ACKNOWLEDGMENTS

REFERENCES

Albert A, Anderson JA (1984). On the existence of maximum likelihood estimates in logistic regression models. Biometrika *71*, 1–10.

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Anderson C, Brown CE (2010). The functions and dysfunctions of hierarchy. Res Organ Behav *30*, 55–89.

Anderson DR (2008). Model Based Inference in the Life Sciences: A Primer on Evidence, New York: Springer.

Andrews TM, Leonard MJ, Colgrove Ca, Kalinowski ST (2011). Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ *10*, 394–405.

Barton K (2015). MuMIn: Multi-Model Inference. R package version 1.15.1. http://CRAN.R-project.org/package=MuMIn.

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

Beichner RJ, Saul JM, Abbott DS, Morse JJ, Duane L, Allain RJ, Bonham SW, Dancy MH, Risley JS (2007). The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) Project. Rev Phys Educ Res *1*, 1–42.

Borrego M, Cutler S, Prince M, Henderson C, Froyd JE (2013). Fidelity of implementation of research-based instructional strategies in engineering science courses. J Eng Educ *102*, 394–425.

Burnham KP, Anderson DR (2003). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, New York: Springer.

Chi MTH, Wylie R (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. Educ Psychol *49*, 219–243.

Cohen EG, Lotan RA, Scarloss BA, Adele R (1999). Complex instruction: equity in cooperative learning classrooms. Theory Pract *38*, 80–86.

Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. Am J Phys *69*, 970–977.

Csernica J, Hanyka M, Hyde D, Shooter S, Toole M, Vigeant M (2002). Practical Guide to Teamwork, Lewisberg, PA: College of Engineering, Bucknell University.

Eddy SL, Brownell SE (2016). Beneath the numbers: a review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. Phys Rev ST Phys Educ Res: Gender in Physics (*in press*).

Eddy SL, Brownell SE, Wenderoth MP (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. CBE Life Sci Educ *13*, 478–492.

Eddy SL, Hogan KA (2014). Getting under the hood: how and for whom does increasing course structure work? CBE Life Sci Educ *13*, 453–468.

Fox J, Weisberg HS (2010). An R Companion to Applied Regression, Thousand Oaks, CA: Sage.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci USA *111*, 8410–8415.

Fritschner LM (2000). Inside the undergraduate college classroom: faculty and students differ on the meaning of student participation. J Higher Educ *71*, 342–362.

Garamszegi LZ (2011). Information-theoretic approaches to statistical analysis in behavioural ecology: an introduction. Behav Ecol Sociobiol *65*, 1–11.

Howard J, Zoeller A, Pratt Y (2006). Students' race and participation in sociology classroom discussion: a preliminary investigation. J Scholarship Teach Learn *5*, 14–38.

Hyde CA, Ruth BJ (2002). Multicultural content and class participation: do students self-disclose? J Soc Work Educ *38*, 241–256.

Jacobs GM, Hall S (2002). Implementing cooperative learning. In: An Anthropology of Current Practice, ed. JC Richards and WA Renandya, Cambridge, UK: Cambridge University Press, 52–58.

James MC, Barbieri F, Garcia P (2008). What are they talking about? Lessons learned from a study of peer instruction. Astron Educ Rev *7*, 37–42.

James MC, Willoughby S (2011). Listening to student conversations during clicker questions: what you have not heard might surprise you! Am J Phys *79*, 123.

Jensen JL, Lawson A (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. CBE Life Sci Educ *10*, 64–73.

Johnson DW, Johnson RT, Smith KA (1998). Cooperative learning returns to college: what evidence is there that it works? Change *30*, 26–35.

Johnson MA, Lawson AE (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? J Res Sci Teach *35*, 89–103.

Karp DA, Yoels WC (1976). The college classroom: some observations on the meanings of student participation. Sociol Soc Res *60*, 421–439.

Kelsey BL (2000). Increasing minority group participation and influence using a group support system. Can J Adm Sci *17*, 63–75.

Kim HS (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. J Pers Soc Psychol *83*, 828–842.

Kim HS (2008). Culture and the cognitive and neuroendocrine responses to speech. J Pers Soc Psychol *94*, 32–47.

Knight JK, Wise SB, Southard KM (2013). Understanding clicker discussions: student reasoning and the impact of instructional cues. CBE Life Sci Educ *12*, 645–654.

Koenig R (2015). In STEM courses, a gender gap in online class discussions. Chron High Educ, January 6.

Kotsopoulos D (2010). When collaborative is not collaborative: supporting student learning through self-surveillance. Int J Educ Res *49*, 129–140.

Ladson-Billings G (2006). From achievement gap to education debt: understanding achievement in U.S. schools. Educ Researcher *35*, 3–12.

Meece JL, Anderman EM, Anderman LH (2006). Classroom goal structure, student motivation, and academic achievement. Annu Rev Psychol 57, 487–503.

Menekse M, Stump GS, Krause S, Chi MTH (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. J Eng Educ 102, 346–374.

Micari M, Drane D (2011). Intimidation in small learning groups: the roles of social-comparison concern, comfort, and individual characteristics in student academic outcomes. Act Learn High Educ 12, 175–187.

Morris LC (2007). Power and status in small groups: an analysis of students' verbal and nonverbal behavior and responses to one another, PhD thesis, University of North Carolina–Chapel Hill.

Ocker R (2007). A balancing act: the interplay of status effects on dominance in virtual teams. Prof Commun IEEE Trans 50, 204–218.

Oliver R (1974). Expectancy theory predictions of salesmen's performance. J Marketing Res 11, 243–253.

Ong M (2005). Body projects of young women of color in physics: intersections of gender, race, and science. Soc Probl 52, 593–617.

Peters RA (1978). Effect of anxiety, curiosity, and perceived instructor threat on student verbal behavior in the college classroom. J Educ Psychol 70, 388–395.

Robnett R (2013). The role of peer support for girls and women in the STEM pipeline: implications for identity and anticipated retention. Int J Gender Sci Technol 5, 233–253.

Roseth CJ, Garfield JB, Ben-Zvi D (2008). Collaboration in learning and teaching statistics. J Stat Educ 16, 1.

Singer SR, Nielsen NR, Schweingruber Ha (2013). Biology education research: lessons and future directions. CBE Life Sci Educ 12, 129–132.

Smith M, Wood W, Adams W (2009). Why peer discussion improves student performance on in-class concept questions. Science 323, 122–124.

Springer L, Stanne ME, Donovan SS (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: a meta-analysis. Rev Educ Res 69, 21–51.

Strough J, Berg CA, Meegan SP (2001). Friendship and gender differences in task and social interpretations of peer collaborative problem solving. Soc Dev 10, 1–22.

Stryker S (2008). Transgender history, homonormativity, and disciplinarity. Radical Hist Rev 100, 145–157.

Tanner KD (2013). Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. CBE Life Sci Educ 12, 322–331.

Tanner KD, Chatman LC, Allen DE (2003). Cooperative learning in the science classroom: beyond students working in groups. Cell Biol Educ 2, 1–5.

Thompson M, Sekaquaptewa D (2002). When being different is detrimental: solo status and the performance of women and racial minorities. Anal Soc Issues Public Policy 2, 183–203.

Venerables WN, Ripley BD (2002). Modern Applied Statistics with S, New York: Springer.

Walton GM, Cohen GL (2007). A question of belonging: race, social fit, and achievement. J Pers Soc Psychol 92, 82–96.

Walton GM, Cohen GL, Cwir D, Spencer SJ (2012). Mere belonging: the power of social connections. J Pers Soc Psychol 102, 513–532.

Weaver RR, Qi J (2005). Classroom organization and participation: college students' perceptions. J Higher Educ 76, 570–601.

White JW (2011). Resistance to classroom participation: minority students, academic discourse, cultural conflicts, and issues of representation in whole class discussions. J Lang Identity Educ 10, 250–265.

Wickett MS (2000). Nurturing the voices of young mathematicians with dyads and group discussions. Teaching Children Mathematics 6.6, 412.

Willoughby T, Wood E, Mcdermott C, Mclaren J (2000). Enhancing learning through strategy instruction and group interaction: is active generation of elaborations critical? 14, 19–30.

Zuur A (2009). Mixed Effects Models and Extension in Ecology with R, New York: Springer.