Gender Differences in Student Participation in an Active-Learning Classroom

Stepfanie M. Aguillon,[†‡§]* Gregor-Fausto Siegmund,[†§]* Renee H. Petipas,[†‖] Abby Grace Drake,[†] Sehoya Cotner,[¶] and Cissy J. Ballen[‖¶#]

[†]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; [‡]Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850; [‖]Plant Pathology Department, Washington State University, Pullman, WA 99163; [¶]Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN 55108; [#]Department of Biological Sciences, Auburn University, Auburn, AL 36849

ABSTRACT

Overwhelming evidence demonstrating the benefits of active-learning pedagogy has led to a shift in teaching that requires students to interact more in the classroom. To date, few studies have assessed whether there are gender-specific differences in participation in active-learning science, technology, engineering, and mathematics (STEM) courses, and fewer have looked across different types of classroom participation. Over two semesters, we observed an introductory biology course at a large research-intensive university and categorized student participation into seven distinct categories to identify gender gaps in participation. Additionally, we collected student grades and administered a postcourse survey that gauged student scientific self-efficacy and salience of gender identity. We found that men participated more than expected based on the class composition in most participation categories. In particular, men were strongly overrepresented in voluntary responses after small-group discussions across both semesters. Women in the course reported lower scientific self-efficacy and greater salience of gender identity. Our results suggest that active learning in itself is not a panacea for STEM equity; rather, to maximize the benefits of active-learning pedagogy, instructors should make a concerted effort to use teaching strategies that are inclusive and encourage equitable participation by all students.

INTRODUCTION

Efforts to identify effective and inclusive teaching methods in science, technology, engineering, and mathematics (STEM) education have led to the adoption of active-learning pedagogy in undergraduate STEM courses (e.g., Smith *et al.*, 2009; Haak *et al.*, 2011; Freeman *et al.*, 2014). The increased opportunity for peer discussions in active-learning classrooms improves student performance (Smith *et al.*, 2009), and the increased structure closes performance gaps (Haak *et al.*, 2011). Many studies show that active learning has a disproportionate benefit on the performance of groups historically underrepresented in STEM (Lorenzo *et al.*, 2006; Eddy and Hogan, 2014; Ballen *et al.*, 2017b; Casper *et al.*, 2019). In these examples, research has largely focused on the impacts of active learning on student performance outcomes (i.e., grades). However, other student outcomes are also worthy of careful consideration in active-learning environments. Fostering participation in the classroom independent of student grades is important for student development, even when students are already earning good grades. Student participation has been linked to a variety of outcomes that can contribute to academic success: critical-thinking skills (Tsui, 2002), a sense of belonging (Eddy and Hogan, 2014), and decreased anxiety about and better understanding of course content (Fassinger, 2000; Haak *et al.*, 2011; Eddy and Hogan, 2014; Freeman *et al.*, 2014). While student participation is highly encouraged in

Sarah L. Eddy, *Monitoring Editor*

Submitted Mar 22, 2019; Revised Feb 3, 2020; Accepted Feb 19, 2020

CBE Life Sci Educ June 1, 2020 19:ar12

DOI:10.1187/cbe.19-03-0048

[§]These authors contributed equally to this work and should be regarded as joint first authors.

*Address correspondence to: Stepfanie M. Aguillon (sma256@cornell.edu) or Gregor-Fausto Siegmund (gs589@cornell.edu).

active-learning classrooms (Handelsman *et al.*, 2004), strategies that promote or inhibit these interactions remain poorly understood.

Activities that encourage participation in the classroom do not always result in positive student outcomes (e.g., Eddy and Hogan, 2014; Knight *et al.*, 2016; Ballen *et al.*, 2017a; Cooper *et al.*, 2018a). In particular, individual beliefs and emotional states influence how comfortable students are in classrooms that encourage participation (Cohen *et al.*, 2019). For instance, self-efficacy (Galyon *et al.*, 2012) and sense of social belonging (Wilson *et al.*, 2015) have been shown to influence student participation. Understanding student affect can help contextualize the outcomes of participation that occur in active-learning classrooms.

Affective measures in active-learning environments can be influenced by axes of identity that are negatively stereotyped, marginalized, or underrepresented (Cooper and Brownell, 2016; Henning *et al.*, 2019). Here, we focus on gender identity and its impacts on student affect in classrooms. For example, scientific self-efficacy is often higher in men than women in STEM (reviewed in Eddy and Brownell, 2016), and women often report a higher salience of gender identity (Picho and Brown, 2011; Cokley *et al.*, 2015). These differences may help explain observations of gender disparities in classroom participation. One of the few studies that has quantified student participation found that women participated less than expected across 23 introductory biology courses (Eddy *et al.*, 2014). Eddy *et al.* (2014) found that women made up on average 60% of the students in the courses, but their voices were only heard 40% of the time in response to instructor questions. More recently, Ballen *et al.* (2019) showed equitable participation between men and women occurred in STEM courses at six institutions in smaller classes and when instructors used diverse teaching strategies. Other studies have demonstrated the pervasiveness of differences in participation between men and women in multiple disciplines (Crombie *et al.*, 2003; Tatum *et al.*, 2013; Ballen *et al.*, 2018; Neill *et al.*, 2018) and at other stages of academic development (e.g., at seminars and conferences; Pritchard *et al.*, 2014; Hinsley *et al.*, 2017; Schmidt and Davenport, 2017; Schmidt *et al.*, 2017; Carter *et al.*, 2018). Because the experiences of undergraduate students in the classroom are consequential (e.g., Tsui, 2002; Haak *et al.*, 2011; Eddy and Hogan, 2014; Freeman *et al.*, 2014), exploring these individual experiences is the first step toward creating equitable learning environments for all students.

Here, we consider gender differences in classroom participation across multiple semesters in a large, introductory biology course. We break down classroom participation into seven categories (descriptions in *Methods: Observational Methods* and Table 1), including unprompted responses, various types of voluntary and nonvoluntary prompted responses, and group work, to understand how the types of interactions that instructors foster in the classroom influence student participation. To gain a holistic understanding about the experiences of men and women in the course, we combined the observations of classroom participation with separately administered surveys that assess salience of gender identity, or the extent to which one's gender forms a central part of one's self-concept in scientific contexts (Picho and Brown, 2011; Lane *et al.*, 2012); scientific self-efficacy, or students' perceptions of their own capability to undertake science tasks (Bandura, 1977); and student grades both for exams and non–exam assignments. We selected these additional measures to understand students' decisions to participate in class. In this study, we address two primary questions: (1) Does the use of active-learning pedagogy in this course result in equitable in-class participation for men and women across different types of in-class interactions? (2) Do we observe gender gaps in self-reported scientific self-efficacy, the salience of gender identity, or performance metrics?

METHODS

Course Overview

This study focused on an introductory evolutionary biology and biodiversity course that is required for all biology majors at a large, selective, research-intensive institution in the northeast United States. Students typically enroll in this course during their first year. The course includes three 50-minute lecture sessions and one 50-minute discussion section (led by graduate student teaching assistants) per week. In 2015, the structure of the course changed to incorporate active-learning principles such as prelecture assignments, prelecture quizzes, assigned in-class groups and group work, frequent student participation, iClicker questions, and a redistribution of grading to reward group work and non–exam assignments (for details, see Ballen *et al.*, 2017b). The course is team taught, and lecture content and slides are similar across years (e.g., biodiversity, adaptation, phylogenetics, population genetics, macroevolution, human evolution). Despite the general active-learning framework of the course, instructors varied in how they used active-learning techniques.

Here, we focus on the Fall 2016 and Fall 2017 semesters of the course, which enrolled 244 and 265 students, respectively. A team of three men and two women taught during the 2016 semester, while a team of three men and three women taught during the 2017 semester. Each lecture session is led by one faculty instructor, who lectures, facilitates discussion, and poses questions to students. Students in this course are assigned to a group before the first class and sit with their groups during every lecture session throughout the semester. In 2016, students were assigned to groups randomly. In 2017, students were assigned to groups to maximize. diversity of student backgrounds. From postcourse surveys, we estimated the gender composition of the class to be 55.9% and 57.1% women in 2016 and 2017, respectively.

Observational Methods

Over the course of two semesters, we observed 40 lecture sessions by eight faculty instructors (Supplemental Table S1; minimum: two lectures; maximum: nine lectures; mean: five lectures/instructor) to document classroom participation. During each lecture, a pair of observers sat at different positions in the classroom and independently observed student participation during the entirety of the lecture session. Each time a student interacted with the instructor (see details on interaction types below), the observer recorded (1) the type of interaction and (2) the perceived gender identity of the student. We recorded gender identity as binary by noting whether we perceived the student to be a man or woman based on auditory and visual signals. We may misattribute gender identity based on biases in how we personally perceive a student's gender presentation.

TABLE 1. Categories used to describe interactions between students and instructors

Category	Interaction type	Explanation
Unprompted	Comment	Student makes an unprompted comment.
Unprompted	Question	Student asks an unprompted question or is only generally prompted.
Prompted and voluntary	Individual	Instructor poses a question and a student volunteers.
Prompted and voluntary	Post-discussion	Instructor poses a question, allows for peer discussion, and a student volunteers.
Prompted and voluntary	Post-clicker	Instructor poses a question, allows for peer discussion, students commit to an answer, and a student volunteers.
Group random call	Group random call	Instructor poses a question, allows for peer discussion, and randomly calls on a group.
Group work	Group work	Student talks directly with the instructor during peer discussion or activity.

Although gender identity is not a binary (Richards *et al.*, 2016), here we code gender as such, because we were unable to match our observations of student participation and perceived gender with students' self-reported gender identity.

At the end of each lecture, the in-class observers pooled observation data to create a single data set. In situations in which observations were missing or did not match between observers, we used the observations from the observer who had a significantly better vantage point of a particular interaction. We did not record gender identity when observers did not agree on the perceived gender of the student or when no observer was able to see a student during an interaction. We attended sessions in person, because it would have been difficult to capture all interactions in the classroom from video recordings alone. The instructors knew that we were observing the classroom, but were not made aware of the particular days we would observe or that our observations focused on classroom participation. Due to the large number of students in the course and the auditorium-style room, we do not believe our presence disrupted or influenced student participation or the instructor teaching the course.

We extended the categories developed in Eddy *et al.* (2014) based on pilot observations conducted in 2016 to more finely evaluate interactions between students and the instructor (Table 1; Ballen *et al.*, 2017a, 2018). We classified interactions into seven distinct types that we pooled into four broad categories: unprompted, prompted and voluntary, group random call, and group work. The four broad categories reflect different teaching strategies that instructors use to elicit participation. *Unprompted interactions* were those that occurred without prompting from the instructor and include (1) comments from students ("comment") and (2) questions from students ("question"). "Question" also includes interactions in which students were generally prompted (e.g., the instructor asked, "Any questions?"). *Prompted and voluntary interactions* occurred when an instructor posed a question to the class and asked for a volunteer to answer: (3) individually without peer discussion ("individual"), (4) after students discussed the question in peer groups ("post-discussion"), and (5) after students discussed the question in peer groups and committed to an answer ("post-clicker"). We refer to this last category as "post-clicker," because students often committed to an answer using iClickers. These three interaction types represent teaching practices with increasing levels of structure (e.g., a post-discussion interaction would become a post-clicker interaction if students were first asked to commit to an answer). *Group random call interactions* were prompted, but nonvoluntary. These interactions (6) occurred when an instructor posed a question to the class, allowed time for discussion in peer groups, and then randomly called on a group to answer ("group random call"). Because whole groups were called on, group members decided which student from the group would report back to the whole class. *Group work interactions* are the only interaction type in which student participation did not take place in front of the entire class. These interactions (7) occurred directly between the instructor and an individual student during peer discussion or small-group activities ("group work"). We could not differentiate interactions initiated by the instructor from those initiated by the student. We recorded only the first interaction an instructor had with a student in a group, even if the instructor spoke to multiple students in the group.

In 2016, we observed 205 student–instructor interactions in 18 lecture sessions taught by four instructors. In 2017, we observed 230 student–instructor interactions in 22 lecture sessions taught by six instructors. We focus our analysis on the student–instructor interactions for which we recorded the student's gender identity as a man or woman (414 out of a total of 435 observations). The data set we analyze consists of 196 observations in 2016 and 218 observations in 2017. Each 50-minute lecture session was structured as an active-learning lecture, with at least 20 minutes of student activities and many opportunities for participation. Students also participated in weekly discussion sections guided by teaching assistants that were not observed for the purposes of this study. In 2016, two observers (SMA, GS) conducted observations; both observers were present at 12/18 lectures, and each observer conducted observations at an additional three lectures independently. In 2017, three observers (SMA, GS, RHP) rotated randomly so that at least two were present at each observation. The observers personally identify such that, in 2016, the sessions were observed by one woman and one man, and in 2017, six out of 22 sessions were observed by two women, and the remaining sessions by one woman and one man. To distribute observation sessions across instructors and course topics, we observed at least two randomly chosen lectures per instructor.

Statistical Analyses of Student Participation

We ran separate analyses for each type of student–instructor interaction to determine whether there was a gender difference in student participation. In each year, we pooled data from all instructors and lecture sessions. To compare the observed counts of interactions by men and women with the expected counts (based on the estimated gender composition of the class), we performed Pearson's χ^2 tests using the chisq.test function in the stats package in R (R Core Team, 2018). In 2017, we had enough observations of lecture sessions taught by men and women to additionally assess whether instructor gender influenced student participation. We ran separate analyses for pooled interaction categories (Table 1; unprompted, prompted and voluntary, group random call, and group work) to determine whether there was a gender difference in student participation that varied with instructor gender. We then performed Pearson's χ^2 tests to compare counts of observed interactions by men and women with the expected counts (based on the estimated gender composition of the class). A conservative assessment of statistical significance would include a Bonferroni correction that adjusts the significance level (α). In the *Results*, we report unadjusted p-values in the text and tables, but report the adjusted significance level in the table legends.

Statistical Analyses of Student Attitudes

At the end of each semester, students were emailed a request to complete a survey that included questions on a variety of topics, from scientific self-efficacy to test anxiety and study habits. As these surveys were not designed specifically for this study, we analyzed the blocks of questions on scientific self-efficacy and the salience of gender identity (Table 2). The questions on scientific self-efficacy have been used and validated in other contexts (Walker *et al.*, 2008; Cotner *et al.*, 2011, 2017; Ballen *et al.*, 2017b), and the questions on salience of gender identity are from a construct within a larger measure called the Social Identities and Attitudes Scale (Picho and Brown, 2011; Smith and Cokley, 2016). We coded responses numerically; responses to self-efficacy questions were on a five-point Likert scale (1 = not confident, 5 = extremely confident), and responses to salience of gender identity questions were on seven-point Likert scale (1 = strongly disagree, 7 = strongly agree).

We analyzed a subset of the data set that included only students who responded to the focal survey questions and identified their gender. We validated the survey structure with a confirmatory factor analysis (CFA) using the cfa function in the lavaan package in R (Rosseel, 2012) following Knekta *et al.* (2019). The specified two-factor CFA demonstrated reasonable model fit for both years (2016: $\chi^2 = 301.439$, $df = 118$, $p < 0.0001$, CFI = 0.907, RMSEA = 0.100, SRMR = 0.053; 2017: $\chi^2 = 324.768$, $df = 118$, $p < 0.0001$, CFI = 0.891, RMSEA = 0.102, SRMR = 0.080), which was further improved by incorporating the covariance between questions ID3 and ID6 (2016: $\chi^2 = 245.701$, $df = 117$, $p < 0.0001$, CFI = 0.936, RMSEA = 0.084, SRMR = 0.047; 2017: $\chi^2 = 206.998$, $df = 117$, $p < 0.0001$, CFI = 0.953, RMSEA = 0.067, SRMR = 0.069). Good model fit values are a comparative fit index (CFI) > 0.95 in combination with root-mean-square error of approximation (RMSEA) < 0.06 or a standardized root-mean-square residual (SRMR) < 0.08 (Hu and Bentler, 1999). Because the CFA model fit was acceptable, we averaged the survey responses within each factor for each student, and then compared responses of women and men using Wilcoxon rank-sum tests with the wilcox.test function in the stats package in R (R Core Team, 2018).

Statistical Analyses of Course Grades

We combined student-reported gender identities with various aspects of course grades to determine whether grades differed based on student gender. Following a previous study that found differences based on type of assessment (Cotner and Ballen, 2017), we separately compared (1) pooled exam grades, (2) pooled non–exam grades (e.g., discussion sections, in-class activities, prelecture assignments), and (3) final course grades of men and women using Wilcoxon rank-sum tests with the wilcox.test function in the stats package in R (R Core Team, 2018). In both semesters, pooled exam grades and pooled non–exam grades comprised 42% and 58% of the final grades, respectively (though the specific non–exam assignments differed between the two semesters). For consistency, we removed students who were enrolled in a writing-intensive discussion section, as their overall course grading schemes differed from those for the rest of the students (nine students in 2016, 13 students in 2017). For ease of comparison, we present grades as transformed z-scores, a measure of how many standard deviations a value is

TABLE 2. Postcourse survey questions focused on scientific self-efficacy and salience of gender identity

Scientific self-efficacy questions	
SE1	Presently I am confident that I can discuss scientific concepts with my friends or family.
SE2	Presently I am confident that I can think critically about scientific findings I read about in the media.
SE3	Presently I am confident that I can read the scientific literature (current papers written by scientists in scientific journals).
SE4	Presently I am confident that I can determine what is and is not valid scientific evidence.
SE5	Presently I am confident that I can make an argument using scientific evidence.
SE6	Presently I am confident that I can present scientific results in writing or orally.
SE7	Presently I am confident that I can understand scientific processes behind important scientific issues in the media.
SE8	Presently I am confident that I can understand the science content of this course.
SE9	Presently I am confident that I can use scientific thinking to solve problems outside this course.
Salience of gender identity questions	
ID1	My gender influences how I feel about myself.
ID2	My gender affects how people treat me.
ID3	My gender is central in defining who I am.
ID4	Most people judge me on the basis of my gender.
ID5	My gender affects how people act toward me.
ID6	My identity is strongly tied to my gender.
ID7	My gender influences how teachers interpret my behavior.
ID8	People of the opposite sex interpret my behavior based on my gender.

from the mean (i.e., negative values indicate students performed below the class mean, positive values indicate students performed above the class mean).

RESULTS

Patterns of Student Participation in an Active-Learning Course

Across 40 lectures by eight instructors over two semesters, we observed variation in the quantity and type of interactions that occurred in the classroom (Figure 1). The two most frequent interaction types in 2016 were group random call (27.0%) and post-discussion prompted and voluntary interactions (19.9%), and in 2017 were group work (33.9%) and group random call (23.4%). The least frequent interaction types in 2016 were unprompted comments (4.6%) and unprompted questions (5.6%), and in 2017 were unprompted comments (1.8%) and post-clicker prompted and voluntary interactions (3.7%).

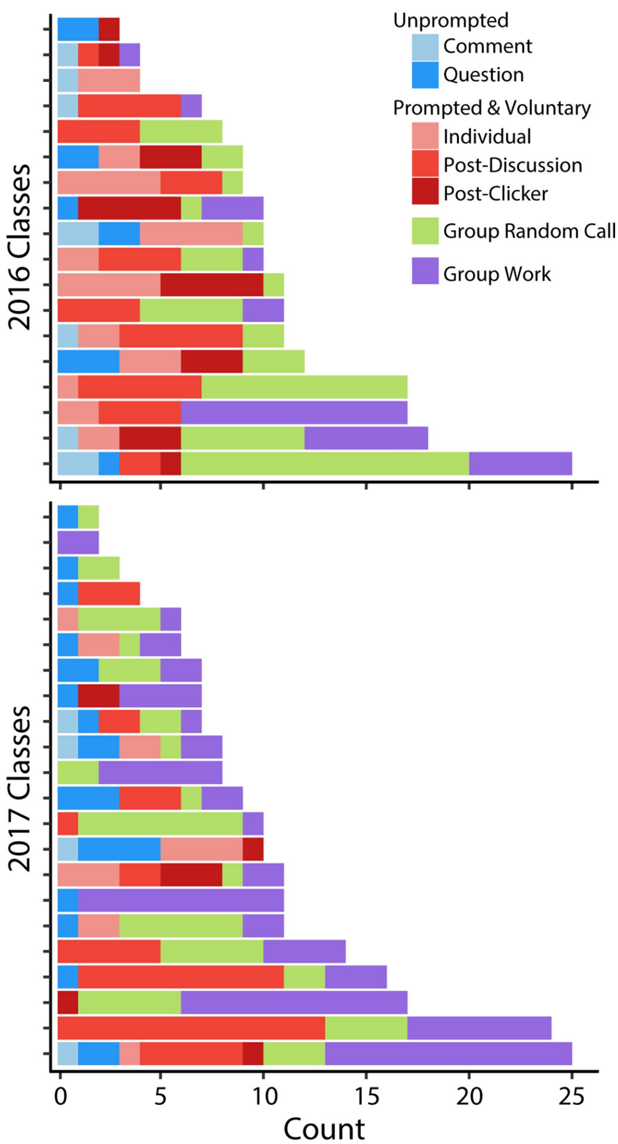

FIGURE 1. Variation in number and type (color of bars) of student participation across observed lecture sessions. See Table 1 for details on different interaction categories. Lecture sessions are ordered by total number of interactions.

In both years, men participated more than expected in post-discussion interactions (Figure 2 and Table 3; 2016: $\chi^2 = 10.003$, $p = 0.002$; 2017: $\chi^2 = 9.488$, $p = 0.002$). In 2016, men participated more than expected in post-clicker interactions (Figure 2 and Table 3; $\chi^2 = 7.321$, $p = 0.007$). In 2017, men participated more than expected in unprompted questions ($\chi^2 = 10.594$, $p = 0.001$), group random call ($\chi^2 = 5.262$, $p = 0.022$), individual prompted and voluntary interactions ($\chi^2 = 5.662$, $p = 0.017$), and group work (Figure 2 and Table 3; $\chi^2 = 6.963$, $p = 0.008$). We do not present statistical results for the interaction types where the expected number of interactions for men or women is less than five (i.e., unprompted comment and unprompted question in 2016 and unprompted comment and post-clicker prompted and voluntary in 2017).

In 2017, we had sufficient observations to assess whether patterns of student participation differed with the gender of the instructor (Table 4). Regardless of instructor gender, men participated more than expected in prompted and voluntary interactions (women instructors: $\chi^2 = 14.688$, $p = 0.0001$; men instructors: $\chi^2 = 8.416$, $p = 0.004$). When the instructor was a woman, men participated more than expected through unprompted interactions ($\chi^2 = 15.522$, $p = 0.0001$), group random call ($\chi^2 = 5.778$, $p = 0.016$), and group work ($\chi^2 = 5.881$, $p = 0.015$). We do not present statistical results for unprompted interactions when the instructor was a man, because the expected number of interactions is less than five.

Course Survey Responses and Performance Metrics

Of the students, 69.1% and 75.9% responded to the postsemester survey and indicated their gender in 2016 ($n = 185$) and 2017 ($n = 183$), respectively. We found significant statistical differences between the responses of men and women to the questions about the salience of gender identity in both years (Figure 3A; 2016: Wilcoxon statistic = $W = 2875$, $p = 0.0006$; 2017: $W = 2645.5$, $p = 0.0025$). We did not find

TABLE 3. Pearson's χ^2 tests for observations of each interaction type pooled by year

Interaction type	*df*	χ^2	p^a
2016 Observations			
Comment	–	–	–
Question	–	–	–
Individual	1	1.061	0.303
Post-discussion	1	10.003	0.002
Post-clicker	1	7.321	0.007
Group random call	1	1.644	0.200
Group work	1	3.693	0.055
2017 Observations			
Comment	–	–	–
Question	1	10.594	0.001
Individual	1	5.662	0.017
Post-discussion	1	9.488	0.002
Post-clicker	–	–	–
Group random call	1	5.262	0.022
Group work	1	6.963	0.008

[a]Adjusted significance level ($\alpha = 0.05/7 = 0.0071$).

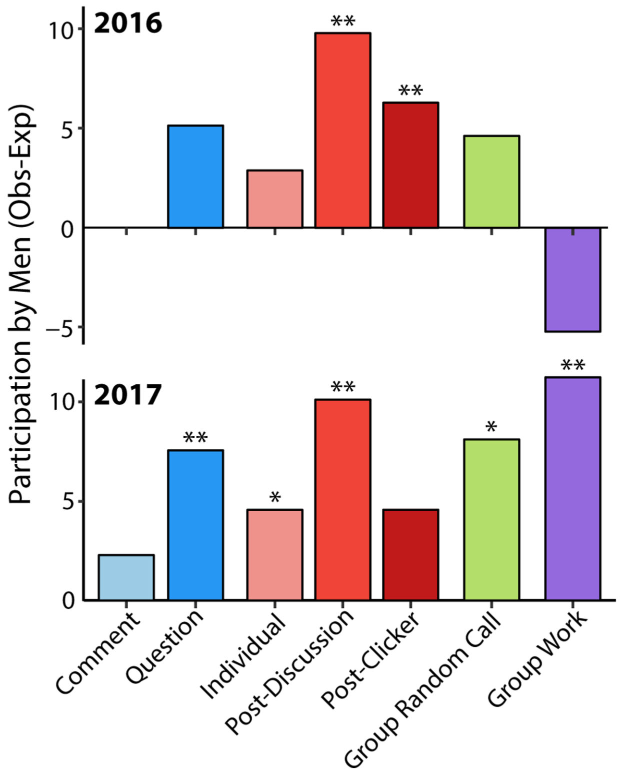

FIGURE 2. In both years, we observed more interactions between instructors and men than expected based on the gender composition of the course across most interaction types (* $p < 0.05$, ** $p < 0.01$). The *y*-axis depicts the difference between the number of interactions by men that were observed and the number of interactions that were expected: positive values indicate more participation by men than expected, while negative values indicate more participation by women than expected. Expected numbers of interactions were calculated as the product of the total number of interactions in a category and the proportion of men in the class. See Table 3 for the corresponding χ^2 tests. Note that sample sizes were not large enough to perform tests on unprompted comment (2016 and 2017), unprompted question (2016), and prompted and voluntary post-clicker (2017).

differences between the responses of men and women to the scientific self-efficacy questions in 2016 (Figure 3B; $W = 4497.5$, $p = 0.4295$), but men reported significantly higher scientific self-efficacy in 2017 (Figure 3B; $W = 4665$, $p = 0.0112$).

Across both semesters, men had significantly higher exam grades than women (2016: $W = 4722.5$, $p = 0.0074$; 2017: $W = 5176$, $p = 0.0003$). However, there was no difference in grades for non–exam assignments in either semester (2016: $W = 4041$, $p = 0.5152$; 2017: $W = 4490$, $p = 0.1091$). The strong difference in exam grades translates into marginally and significantly higher final grades for men in 2016 ($W = 4400$, $p = 0.0855$) and 2017 ($W = 4990$, $p = 0.0023$), respectively.

DISCUSSION

When students are exposed to active-learning pedagogy and tasked with interacting with their peers and instructor in a class, they tend to achieve more (Smith *et al.*, 2009), understand better (Haak *et al.*, 2011; Eddy and Hogan, 2014; Freeman *et al.*, 2014), hold a higher sense of belonging (Eddy and Hogan, 2014), and have higher science self-efficacy (Ballen *et al.*, 2017b) than students in lecture-based classrooms. Studies that have demonstrated that women participate less than expected in college STEM classrooms (Eddy *et al.*, 2014) underscore the need to conduct empirical work to test whether students engage with and benefit from active learning equitably. In this study, we focused on nuanced patterns of student participation, attitudes, and performance in a large, introductory active-learning biology course.

TABLE 4. Pearson's χ^2 tests for observations of interactions in 2017 pooled by interaction category and instructor gender

Instructor gender	*df*	χ^2	p^a
Unprompted			
Women	1	15.522	0.0001
Men	–	–	–
Prompted and voluntary			
Women	1	14.688	0.0001
Men	1	8.416	0.004
Group random call			
Women	1	5.778	0.016
Men	1	0.368	0.544
Group work			
Women	1	5.881	0.015
Men	1	1.190	0.275

[a]Adjusted significance level ($\alpha = 0.05/8 = 0.0063$).

We designed our study to assess student engagement with instructional strategies in an active-learning classroom. We observed variation in the quantity and type of interactions between instructors and students (Figure 1), a majority of which followed active-learning practices (post-discussion, post-clicker, group random call, group work): 73.5% in 2016 and 81.2% in 2017. Before the study, we hypothesized that more men than women would participate when instructors interacted with students through unprompted comments, unprompted questions, and individual prompted and voluntary responses. We hypothesized that we would observe parity in whole-class participation when instructors used active-learning practices such as peer discussion and clicker questions, as well when instructors randomly called on groups. We also hypothesized that we would observe parity in interactions between instructors and students during group work.

In our observations, men participated more than women in all interaction types but one in both semesters of our study (Figure 2), although the statistical significance varied (Table 3). Men participated more than expected, relative to the gender composition of the course, in post-discussion prompted and voluntary interactions in both semesters and in post-clicker prompted and voluntary interactions in 2016. Eddy *et al.* (2014) also found that men participated more through volunteering, though our interaction types are not directly comparable. While that study did not identify differences in participation between men and women in spontaneous questions (Eddy *et al.*, 2014), we observed men participating more in this interaction type in 2017 (the only year we have enough data). In our study, when groups were randomly called on by instructors, there was no difference in participation between men and

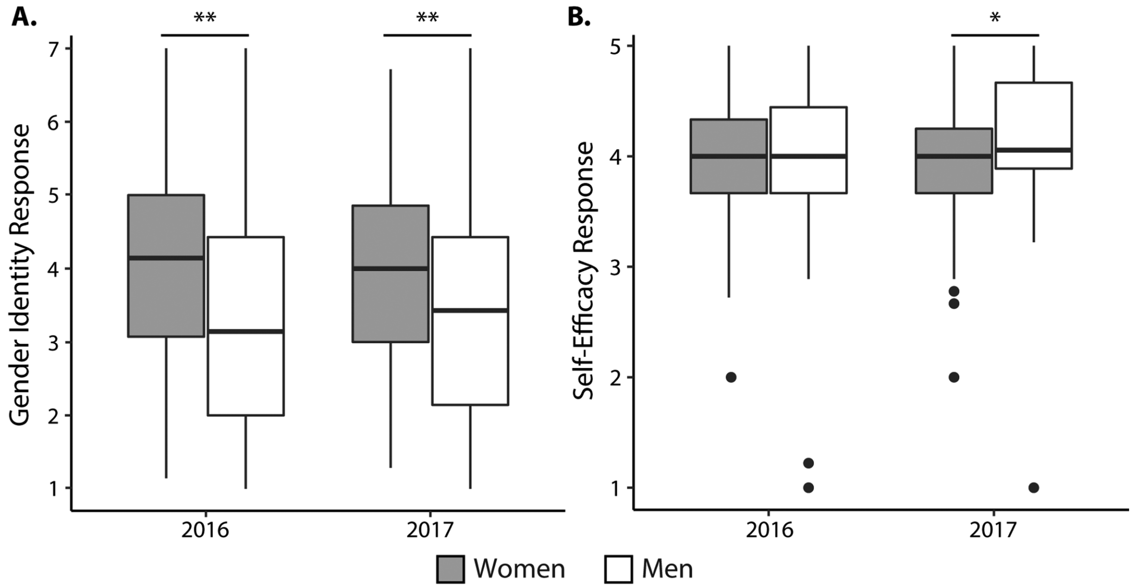

FIGURE 3. Differences in student attitudes based on postcourse surveys (see Table 2 for the list of questions). (A) Women (gray bars) responded significantly more strongly than men (white bars) on gender identity questions across both years. (B) In 2017, men had significantly higher scientific self-efficacy responses than women, but no significant differences were found in 2016 (* $p < 0.05$, ** $p < 0.01$).

women in 2016, but more men participated than expected in 2017. Men also participated more than expected in group work in 2017.

Random calling has been suggested as a tool to counter gender disparities in student participation (e.g., Eddy *et al.*, 2014, demonstrated equitable participation when instructors randomly called on students from a presorted list). The classroom we observed used "group random call," such that instructors randomly called on whole groups instead of individual students. Group members quickly negotiated who would speak after a group was selected (personal observation). One study comparing the effect of groups volunteering a response, versus group random call, found that the latter promoted a higher quantity and quality of peer discussion, likely due to a higher level of accountability (Knight *et al.*, 2016). Despite the differences in who reported to the whole class in these group random call interactions, it is possible that peer discussions were equitable.

Women in class reported that they identified more strongly with their gender than men and that other people (e.g., their peers, their instructors) judge them more on the basis of gender (Figure 3A). These responses suggest that women may be more likely than men to experience stereotype threat, which is fear of confirming a negative stereotype about one's social identity group in competitive and evaluative contexts (Steele, 1997; Beasley and Fischer, 2012). While we did not assess students' perceptions of their peers in our study, research in undergraduate biology classrooms at another institution demonstrates that men underestimate the academic performance of their women peers (Grunspan *et al.*, 2016). Men reported higher scientific self-efficacy than women in 2017 (Figure 3B) and higher exam grades in both years (see *Results*), despite there being no difference between men and women in performance on non–exam assignments in the course (see *Results*). Differences in student confidence and perceptions about being evaluated may thus contribute to students' decisions to assume an explainer/leader role in their groups (Eddy *et al.*, 2015). Although we did not include measures of incoming knowledge when assessing differences in grades, the performance gaps we observe are those that the students experience in the classroom and that contribute to their self-concepts. Previous work in the same course in 2014 and 2015 showed no gender difference when incoming preparation was included in the analysis (Ballen *et al.*, 2017b), in agreement with other research (e.g., Lauer *et al.*, 2013; Salehi *et al.*, 2019). However, some work in undergraduate biology has documented gender gaps on exams even when controlling for academic ability (e.g., Eddy *et al.*, 2014). Differences between studies may be the result of different student populations, the stakes of the exams (i.e., the extent to which they contribute to course grades), or other factors.

Research suggests that same-gender instructor role models can improve confidence (Cotner *et al.*, 2011), achievement (Eddy *et al.*, 2014), and retention in undergraduate majors (Rask and Bailey, 2002). We hypothesized that women students might participate more when instructors were women, leading to parity in participation. However, we observed that men were more likely to volunteer responses to prompted questions regardless of instructor gender and participated more than expected when the instructor was a woman across all other interaction categories (Table 4). Our observations of unprompted interactions are consistent with studies that find men students are more likely to interrupt women instructors (Brooks, 1982). One limitation of our analysis is that, in 2017, we observed more interactions in each category by women than men instructors (unprompted: 18 vs. eight; prompted and voluntary: 52 vs. 15; group random call: 33 vs. 18; group work: 54 vs. 20), which means that the power of our statistical tests was lower for assessing how students interacted with men instructors. Generally, our analysis of this question is necessarily exploratory, as we did not explicitly design our study to test a hypothesis about student and instructor gender.

Caveats and Limitations

We emphasize that "participation" is the product of actions by students and instructors. Our observations reflect who was heard but not whether women or men knew or contributed explanations or answers during peer discussion. Whole-class interactions after peer discussion could measure who claims or receives credit for answers or who represents the group to the class. Women's expertise is underestimated in group tasks (Thomas-Hunt and Phillips, 2004), and research in an undergraduate STEM course found that men are more likely than women to assess themselves as smarter than their peers (Cooper *et al.*, 2018b). Such perceptions could contribute to decisions about who from a group will take credit for an explanation or answer. The interactions we observed could also be the product of decisions made by groups regarding who will represent them to the class, and men may be more likely to volunteer as spokespersons. Interviews with students, observations of student discussions, and surveys could all be useful for parsing the relative importance of these processes.

More broadly, student participation happens in response to actions by instructors during class and to teaching strategies. For example, observations of student participation may reflect who instructors call on or how instructors ask questions. Instructor gender bias (Moss-Racusin *et al.*, 2012) may affect which students they call on, and the instructional practice of providing students with only a few seconds to answer a question rewards students who answer quickly (Rowe, 1974). In our observations, it was uncommon to have multiple students volunteer in response to any given question, and the voluntary responses we observed were typically from the first person to volunteer (personal observation). It is possible that this contributed to our observations of gender bias in participation, but we are unable to parse instructor practices from student behavior in our data set.

In our study, we assume that whole-class participation and self-reported surveys are related to how students experience the classroom. However, we encourage readers to place our results in the context of scholarship in STEM fields that uses methods such as (non)participant observation, interviews, and ethnography. Work that analyzes student narratives about experiences in STEM classrooms and degrees emphasizes the relevance of identity in student interactions with peers and instructors (Johnson, 2007; Gregory, 2015; Cooper and Brownell, 2016; Hughes *et al.*, 2017). Our results are consistent with these findings, but cannot address the experience of individual students.

Focusing on social gender identity without explicitly considering its intersection with other identities means that we are inevitably missing patterns. For example, women students of color in STEM fields experience a "double bind" of sexist and racist expectations, bias, and actions (Ong *et al.*, 2011). Interviews also suggest that active learning increases the importance of LGBTQIA (lesbian, gay, bisexual, transgender, queer, intersex, and asexual) identities in the classroom (Cooper and Brownell, 2016). Our study was not designed to parse multiple, intersecting identity axes, and we acknowledge that gendered patterns are simply one possible outcome of multiple processes. Finally, we were unable to match self-reported gender identity with observations of participation and recorded gender as a binary. Because our observations may misattribute gender, our results could over- or underestimate the relative importance of student gender. Our study also does not reflect the experiences of students who do not identify as binary or present as men or women.

Recommendations for Future Research on Equitable Participation in Active-Learning Classrooms

Active learning emphasizes peer discussion and low-stakes, in-class assessment, often in the form of random call or clicker questions. Our observations of participation and survey results contribute to research on the limitations of implementing active learning without explicitly considering student outcomes such as equitable participation and positive affect. In particular, the benefits of active learning may not be equally shared among students if equitable teaching strategies are not considered and implemented.

Future research will benefit from a closer examination of the impact of course practices and group dynamics on student participation (e.g., Ballen *et al.*, 2019). To examine course practices, one can manipulate the classroom environment and measure the consequences for students (e.g., equitable participation, affective measures). This approach rejects the "student deficit model" as a way to explain gaps in participation and performance, which assumes such gaps are due to inadequacies of students operating in a fair learning environment. Rather, the "course deficit model" encourages instructors to examine aspects of their own teaching environments that might contribute to observed disparities (Cotner and Ballen, 2017). For example, prompted and voluntary post-discussion and post-clicker interactions are in some ways at the core of the active-learning strategies implemented in this course, yet these were also the categories in which men consistently participated more than expected. One can compare the currently reported patterns with those of future courses that employ teaching practices explicitly targeting equitable participation (Tanner, 2013).

Examining group dynamics is another fruitful avenue for research, as group interactions seem to play an important part in active-learning classrooms. Peer discussion among group members encourages students to defend and articulate their ideas, promoting mastery and engagement (Hodges, 2018), particularly when instructors provide structured learning environments. For example, Knight *et al.* (2016) compared the quality of group discussions among those who expect the instructor to ask for a volunteer to respond (less structured), and those who expect they may be randomly called on for an answer (more structured). They found that the groups who thought they might be randomly called on achieved a higher level of exchange of quality reasoning in their discussions. Similarly, activities that include turn-taking and explicit prompts for students to participate improved comfort within groups (Theobald *et al.*, 2017). Future research that focuses explicitly on best practices for group work and cohesion will also enhance whole-class participation. Our observational data set did not allow us to explore how group composition affected participation, but past research has found that the gender ratio of the "microenvironment" created by small groups can affect participation by women (Dasgupta *et al.*, 2015). Continued work on student composition and its effects on learning and group dynamics will also lead to improved climate and student outcomes (Gruenfeld *et al.*, 1996; Theobald *et al.*, 2017; Sullivan *et al.*, 2018). Finally, we recommend future work studying the impacts of educating instructors about the pervasiveness of participation gaps and the benefits of equitable teaching strategies. Research on structure and equity in active-learning classrooms will help instructors identify equitable teaching strategies to benefit future cohorts of undergraduate students.

ACKNOWLEDGMENTS

We thank K. Williams and C. McLinn at the Center for the Integration of Research, Teaching, and Learning at Cornell University for help with the early development of this project, L. Johnson for statistical support, and the many instructors and teaching assistants who help make active learning possible in this course. Members of the Fuller Evolutionary Biology Lab at the Cornell Lab of Ornithology and the DBER Journal Club provided thoughtful comments on this project. The thorough comments from two anonymous reviewers greatly improved this article. This work was approved by Cornell University Institutional Review Board protocol 1410005010.

REFERENCES

Ballen, C. J., Aguillon, S. M., Awwad, A., Bjune, A. E., Challou, D., Drake, A. G., … & Cotner, S. (2019). Smaller classes promote equitable student participation in STEM. *BioScience*, *69*(8), 669–680. doi: 10.1093/biosci/biz069

Ballen, C. J., Danielsen, M., Jørgensen, C., Grytnes, J. A., & Cotner, S. (2017a). Norway's gender gap: Classroom participation in undergraduate introductory science. *Nordic Journal of STEM Education*, *1*(1), 262–270. doi: 10.5324/njsteme.v1i1.2325

Ballen, C. J., Lee, D., Rakner, L., & Cotner, S. (2018). Politics a "chilly" environment for undergraduate women in Norway. *PS: Political Science & Politics*, *51*(3), 653–658. doi: 10.1017/S1049096518000045

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017b). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*(4), 1–6. doi: 10.1187/cbe.16-12-0344

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. doi: 10.1037/0033-295X.84.2.191

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448. doi: 10.1007/s11218-012-9185-3

Brooks, V. R. (1982). Sex differences in student dominance behavior in female and male professors' classrooms. *Sex Roles*, *8*(7), 683–684. doi: 10.1007/BF00287565

Carter, A. J., Croft, A., Lukas, D., & Sandstrom, G. M. (2018). Women's visibility in academic seminars: Women ask fewer questions than men. *PLoS ONE*, *13*(9), 1–22. doi: 10.1371/journal.pone.0202743

Casper, A. M., Eddy, S. L., & Freeman, S. (2019). True grit: Passion and persistence make an innovative course design work. *PLoS Biology*, *17*(7), e3000359. doi: 10.1371/journal.pbio.3000359

Cohen, M., Buzinski, S. G., Armstrong-Carter, E., Clark, J., Buck, B., & Reuman, L. (2019). Think, pair, freeze: The association between social anxiety and student discomfort in the active learning environment. *Scholarship of Teaching and Learning in Psychology*, *5*(4), 265–277. https://doi.org/10.1037/stl0000147

Cokley, K., Awad, G., Smith, L., Jackson, S., Awosogba, O., Hurst, A., … & Roberts, D. (2015). The roles of gender stigma consciousness, imposter phenomenon and academic self-concept in the academic outcomes of women and men. *Sex Roles*, *73*(9–10), 414–426. doi: 10.1007/s11199-015-0516-7

Cooper, K. M., & Brownell, S. E. (2016). Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA Students. *CBE—Life Sciences Education*, *15*(3), 1–19. doi: 10.1187/cbe.16-01-0074

Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018a). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, *5*(23), 1–18. doi: 10.1186/s40594-018-0123-6

Cooper, K. M., Krieg, A., & Brownell, S. E. (2018b). Who perceives they are smarter ? Exploring the influence of student characteristics on student academic self-concept in physiology. *Advances in Physiology Education*, *42*(2), 200–208. doi: 10.1152/advan.00085.2017

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable ? *PLoS ONE*, *12*(12), 1–11. doi: 10.1371/journal.pone.0189610

Cotner, S., Ballen, C., Brooks, D. C., & Moore, R. (2011). Instructor gender and student confidence in the sciences: A need for more role models? *Journal of College Science Teaching*, *40*(5), 96–101. doi: 10.2307/42993884

Cotner, S., Thompson, S., & Wright, R. (2017). Do biology majors really differ from non–STEM majors? *CBE—Life Sciences Education*, *16*(3), ar48. doi: 10.1187/cbe.16-11-0329

Crombie, G., Pyke, S. W., Silverthorn, N., Jones, A., & Piccinin, S. (2003). Students' perceptions of their classroom participation and instructor as a function of gender and context. *Journal of Higher Education*, *74*(1), 51–76. doi: 10.1080/00221546.2003.11777187

Dasgupta, N., Scircle, M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences USA*, *112*(16), 4988–4993. doi: 10.1073/pnas.1422822112

Eddy, S. L., & Brownell, S. E. (2016). Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines. *Physical Review Physics Education Research*, *12*(2), 1–20. doi: 10.1103/PhysRevPhysEducRes.12.020106

Eddy, S. L., Brownell, S. E., Thummaphan, P., Lan, M., & Wenderoth, M. P. (2015). Caution, student experience may vary: Social identities impact a student's experience in peer discussions. *CBE—Life Sciences Education*, *14*(4), 1–17. doi: 10.1187/cbe.15-05-0108

Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE—Life Sciences Education*, *13*(3), 478–492. doi: 10.1187/cbe.13-10-0204

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, *13*(3), 453–468. doi: 10.1187/cbe.14-03-0050

Fassinger, P. A. (2000). How classes influence students' participation in college classrooms. *Journal of Classroom Interaction*, *35*(2), 38–47.

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

Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L., & Williams, R. L. (2012). The relationship of academic self-efficacy to class participation and exam performance. *Social Psychology of Education*, *15*(2), 233–249. doi: 10.1007/s11218-011-9175-x

Gregory, S. L. (2015). *African American female engineering students' persistence in stereotype-threatening environments: A critical race theory perspective (Doctoral dissertation)*. Retrieved from Digital Commons at USU (4260).

Gruenfeld, D., Mannix, E., Phillips, K., & Neale, M. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*(1), 1–15. doi: 10.1006/obhd.1996.0061

Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE*, *11*(2), 1–16. doi: 10.1371/journal.pone.0148405

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216. doi: 10.1126/science.1204820

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., … & Wood, W. B. (2004). Scientific teaching. *Science*, *304*(5670), 521–522. doi: 10.1126/science.1096022

Henning, J. A., Ballen, C. J., Molina, S., & Cotner, S. (2019). Hidden identities shape student perceptions of active learning environments. *Frontiers in Education*, *4*, 129. doi: 10.3389/feduc.2019.00129

Hinsley, A., Sutherland, W. J., & Johnston, A. (2017). Men ask more questions than women at a scientific conference. *PLoS ONE*, *12*(10), 1–14. doi: 10.1371/journal.pone.0185534

Hodges, L. C. (2018). Contemporary issues in group learning in undergraduate science classrooms: A perspective from student engagement. *CBE—Life Sciences Education*, *17*(2), 1–10. doi: 10.1187/cbe.17-11-0239

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

Hughes, C. C., Schilt, K., Gorman, B. K., & Bratter, J. L. (2017). Framing the faculty gender gap: A view from STEM doctoral students. *Gender, Work, and Organization*, *24*(4), 398–416. doi: 10.1111/gwao.12174

Johnson, A. C. (2007). Unintended consequences: How science professors discourage women of color. *Science Education*, *91*(5), 805–821. doi: 10.1002/sce.20208

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

Knight, J. K., Wise, S. B., & Sieke, S. (2016). Group random call can positively affect student in-class clicker discussions. *CBE—Life Sciences Education*, *15*(4), 1–11. doi: 10.1187/cbe.16-02-0109

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, *66*(3–4), 220–234. doi: 10.1007/s11199-011-0036-z

Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaia, M., Christensen, W., & Montplaisir, L. (2013). Stereotyped: Investigating gender in introductory science courses. *CBE—Life Sciences Education*, *12*(1), 30–38. doi: 10.1187/cbe.12-08-0133

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*(2), 118–122. doi: 10.1119/1.2162549

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences USA*, *109*(41), 16474–16479. doi: 10.1073/pnas.1211286109

Neill, C., Cotner, S., Driessen, M., & Ballen, C. J. (2018). Structured learning environments are required to promote equitable participation. *Chemistry Education Research and Practice*, *20*(1), 197–203. doi: 10.1039/C8RP00169C

Ong, M., Wright, C., Espinosa, L. L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review*, *81*(2), 172–209. doi: 10.17763/haer.81.2.t022245n7x4752v2

Picho, K., & Brown, S. W. (2011). Can stereotype threat be measured? A validation of the Social Identities and Attitudes Scale (SIAS). *Journal of Advanced Academics*, *22*(3), 374–411. doi: 10.1177/1932202X1102200302

Pritchard, J., Masters, K., Allen, J., Contenta, F., Huckvale, L., Wilkins, S., & Zocchi, A. (2014). Asking gender questions. *Astronomy & Geophysics*, *55*(6), 6.8–6.12. doi: 10.1093/astrogeo/atu245

Rask, K. N., & Bailey, E. M. (2002). Are faculty role models? Evidence from major choice in an undergraduate institution. *Journal of Economic Education*, *33*(2), 99–124. doi: 10.1080/00220480209596461

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry*, *28*(1), 95–102. doi: 10.3109/09540261.2015.1106446

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi: 10.18637/jss.v048.i02

Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic, and fate control: Part one—wait-time. *Journal of Research in Science Teaching*, *11*(2), 81–94. doi: 10.1002/tea.3660110202

Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., … & Ballen, C. J. (2019). Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. *Frontiers in Education*, *4*, 107. doi: 10.3389/feduc.2019.00107

Schmidt, S. J., & Davenport, J. R. A. (2017). Who asks questions at astronomy meetings? *Nature Astronomy*, *1*(6), 0153. doi: 10.1038/s41550-017-0153

Schmidt, S. J., Douglas, S., Gosnell, N. M., Muirhead, P. S., Booth, R. S., Davenport, J. R. A., & Mace, G. N. (2017). The role of gender in asking questions at Cool Stars 18 and 19. Retrieved from arXiv:1704.05260. doi: 10.5281/zenodo.546881

Smith, L. V., & Cokley, K. (2016). Stereotype threat vulnerability: A psychometric investigation of the social identities and attitudes scale. *Measurement and Evaluation in Counseling and Development*, *49*(2), 145–162. doi: 10.1177/0748175615625752

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323*(5910), 122–124. doi: 10.1126/science.1165919

Steele, C. M. (1997). A threat in the air. *American Psychologist*, *52*(6), 613–629. doi: 10.1037/0003-066X.52.6.613

Sullivan, L. L., Ballen, C. J., & Cotner, S. (2018). Small group gender ratios impact biology class performance and peer evaluations. *PLoS ONE*, *13*(4), 1–14. doi: 10.1371/journal.pone.0195129

Tanner, K. D. (2013). Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE—Life Sciences Education*, *12*(3), 322–331. doi: 10.1187/cbe.13-06-0115

Tatum, H. E., Schwartz, B. M., Schimmoeller, P. A., & Perry, N. (2013). Classroom participation and student-faculty interactions: Does gender matter? *Journal of Higher Education*, *84*(6), 745–768. doi: 10.1080/00221546.2013.11777309

Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS ONE*, *12*(7), 1–16. doi: 10.1371/journal.pone.0181336

Thomas-Hunt, M. C., & Phillips, K. W. (2004). When what you know is not enough: Expertise and gender dynamics in task groups. *Personality and Social Psychology Bulletin*, *30*(12), 1585–1598. doi: 10.1177/0146167204271186

Tsui, L. (2002). Fostering critical thinking through effective pedagogy. *Journal of Higher Education*, *73*(6), 740–763. doi: 10.1080/00221546.2002.11777179

Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: Integrating active learning into a large lecture course. *CBE—Life Sciences Education*, *7*(4), 361–367. doi: 10.1187/cbe.08-02-0004

Wilson, D., Jones, D., Bocell, F., Crawford, J., Kim, M. J., Veilleux, N., … & Plett, M. (2015). Belonging and academic engagement among undergraduate STEM students: A multi-institutional study. *Research in Higher Education*, *56*(7), 750–776. doi: 10.1007/s11162-015-9367-x