Gender Identity and Student Perceptions of Peer Research Aptitude in CUREs and Traditional Laboratory Courses in the Biological Sciences

David Esparza,[†] Aimeé A. Hernández-Gaytan, and Jeffrey T. Olimpo*
Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX 79968

ABSTRACT
While several studies have investigated gender inequities in the social learning environment of biology lecture courses, that same phenomenon remains largely unexplored in biology laboratory contexts. We conducted a mixed methods study to understand the influence of gender on student perceptions of their peers' research aptitude in introductory biology CUREs and traditional laboratory courses. Specifically, students (N = 125) were asked to complete a name generator survey at three time points across the semester. This survey asked students to list the names of peers whom they viewed as "most proficient" in the course investigations and to justify their choice via an open-ended response prompt. Using social network analysis, exponential random graph modeling (ERGM), and thematic analysis, we demonstrate that student gender identity did not influence nomination behaviors in CURE or traditional laboratory courses. However, the ERGMs reveal the presence of a popularity effect in CUREs and demonstrate that mutual nominations were more prevalent in traditional laboratory courses. Our qualitative data further provide insights into the reasons students nominated peers as proficient in CURE and traditional courses.

INTRODUCTION
National efforts to increase the representation and participation of all individuals in science, technology, engineering, and mathematics (STEM) have been ongoing since the 1950s (Puaca, 2013). Despite substantial progress being made toward achieving this goal (Bernard and Cooperdock, 2018), high-quality educational experiences in STEM largely remain inaccessible to individuals who identify with marginalized racial, ethnic, and gender identities (Finley and McNair, 2014; Springer *et al.*, 2018). These disparities in opportunity access can impact how marginalized individuals are perceived by others (e.g., instructors, peers, prospective employers).

As identified above, gender identity is one such factor that can impact an individual's opportunities in STEM. At the college level, research indicates that undergraduate women consistently experience diminished knowledge and affective outcomes across STEM disciplines (Matz *et al.*, 2017; Salehi *et al.*, 2019a). These disparities are not reduced as women progress through their chosen degree programs (Eddy and Brownell, 2016) and may result in these individuals departing from said programs altogether (Beasley and Fischer, 2012). Between the years of 2003 and 2009, for instance, a higher percentage of women (32%) than men (26%) declared but left STEM majors for non-STEM majors (Chen, 2013). National statistics in 2018 and 2019 further indicate that, amongst those who persisted, women were still awarded a smaller percentage (36%) of the approximately 413,000 STEM bachelor's degrees that were conferred as compared with men (64%; U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2020).

In contrast to other STEM fields, women represent over 60% of the student population in biology degree programs, and a majority (62%) of 2018 undergraduate

Rebecca Price, *Monitoring Editor*
Submitted Mar 14, 2022; Revised Aug 28, 2023; Accepted Sep 13, 2023
CBE Life Sci Educ December 1, 2023 22:ar53
DOI:10.1187/cbe.22-03-0054
[†]Current Address: Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14850.
*Address correspondence to: Jeffrey T. Olimpo (jtolimpo@utep.edu).

biology degree recipients were women (Ganley *et al.*, 2018; U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2020). These demographic trends have given rise to the perception that biology has "conquered" gender disparities (Eddy *et al.*, 2014). Yet, recent studies find that the experiences of women in biology degree programs are similar to those shared by women across the undergraduate STEM landscape. Multiple studies indicate, for example, that women earn significantly lower grades on biology exams as compared with men (Eddy *et al.*, 2014; Ballen *et al.*, 2017). These differences in student performance can, in part, be explained by the "chilly" climate of undergraduate biology courses. For instance, Grunspan *et al.* (2016) found that women students' knowledge is often underestimated by the men enrolled in their introductory biology lecture courses.

This culture can impact women students' performance and engagement in undergraduate biology lecture environments. Research indicates that women express significantly greater test anxiety relative to men when completing exams in biology lecture courses (Ballen *et al.*, 2017). Further, women have been shown to participate less than men in both traditional, didactic biology lecture courses (Eddy *et al.*, 2014) as well as biology lecture courses that are designed to support active student engagement (Aguillon *et al.*, 2020). However, research indicates that women students exhibit improved performance (Eddy *et al.*, 2014) and/or participation (Bailey *et al.*, 2020) in biology lecture courses taught by women instructors or that have a higher percentage of women students in attendance. Even then, women remain underrepresented in faculty positions that provide them opportunities to teach undergraduate biology courses (Sheltzer and Smith, 2014).

Frequently, laboratory coursework serves to complement lecture content in the undergraduate biology curriculum. Laboratory courses are often designed to provide students with hands-on scientific experiences, wherein students work in groups to conduct scientific investigations. While these courses promote student development of core competencies in biology, ethnographic research shows that students can conflate "doing science" with "doing gender" by adopting stereotypical gender roles when dividing labor during laboratory assignments (Danielsson and Linder, 2009; Gonsalves *et al.*, 2016). For example, research on college physics (Holmes *et al.*, 2014; Quinn *et al.*, 2020) and high-school biology laboratory courses (Kokott *et al.*, 2018) suggests that men spend more time than women using experimental equipment. Conversely, women take on a disproportionate amount of leadership responsibilities and managerial work as compared with men during laboratory assignments (Doucette *et al.*, 2020). While all of these behaviors are necessary for the completion of the laboratory work, prior research indicates that gender stereotyping can influence the way people estimate the aptitude, achievement, and eventual outcomes of women and men (Deaux, 1984).

Course-based undergraduate research experiences (CUREs) have been posed as a method of laboratory education to promote inclusivity, equity, and diversity in scientific research (Bangera and Brownell, 2014). CUREs are designed to support student-led research, wherein students often formulate their own research questions, design experiments, and collect data (Auchincloss *et al.*, 2014). As opposed to apprenticeship-style research experiences – where a few undergraduate trainees research alongside a senior scientist (e.g., faculty mentor) – CUREs enable instructors to engage entire classes of students in intensive research in a course setting (Wei and Woodin, 2011). While CUREs increase the accessibility of undergraduate research, it remains unclear whether CUREs are effective at fostering an inclusive learning environment that supports equity in STEM. Recent research has found that laboratory courses that support more open, unstructured forms of inquiry – characteristic of CUREs – can reinforce gender disparities by fostering a social environment that encourages students to self-sort into stereotypical gender roles (Doucette *et al.*, 2020). Therefore, it is important to examine the extent to which CUREs (and their traditional laboratory counterparts) function as equitable social learning environments.

RESEARCH QUESTIONS AND HYPOTHESES

Grunspan *et al.* (2016) investigated students' perceptions of their peers' aptitude to determine whether men and women enrolled in undergraduate biology lecture courses possessed gendered biases in who they selected as being knowledgeable. The authors found that men often underestimated the intellect of women, whereas women, overall, found both women and men to be similarly knowledgeable. We used nominations in a similar way to define the influence of gender stereotyping and biases in CUREs and traditional biology laboratory courses. Specifically, we investigated the following research questions:

1. What do student nomination structures look like in CUREs and traditional laboratory courses?
2. Does a student's gender identity influence the probability that they will get nominated as proficient by their peers in CURE and traditional courses?
3. For what reasons do students view their peers as proficient, and are there differences in the way students justify their nominations based on the gender identity of the nominee?

We hypothesized that nominations would be more distributed in CURE laboratory courses and more localized in traditional courses. In other words, we anticipated that traditional laboratory students would likely nominate a select few individuals whom they were assigned to work with throughout the semester. Conversely, we predicted that CURE students would be more likely to nominate a broader array of their peers. This hypothesis is guided by the highly collaborative nature of CUREs (Auchincloss *et al.*, 2014) and the tendency for CURE students to interact with peers across assigned working groups (Esparza *et al.*, 2020). Further, we expected that gender identity would influence nomination probability given prior research on the impact of student gender identity on student behaviors in STEM laboratory courses (e.g., Doucette *et al.*, 2020) and student perceptions in biology lecture courses (Grunspan *et al.*, 2016). In both CURE and traditional laboratory courses, we anticipated that students would cite visible characteristics or behaviors of the nominee (e.g., they frequently ask questions) to justify their nomination choices.

MATERIALS AND METHODS

Participant Recruitment and Selection Process

Participants (N = 125) represented a sample of convenience including students (n = 57) enrolled in three introductory cell

TABLE 1. Demography of CURE and traditional students

Category	CURE % (n = 57)	Traditional % (n = 68)
Gender Identity		
Man	28.07	29.41
Woman	59.65	58.82
Unknown[a]	12.28	11.77
Race/Ethnicity		
White	8.77	5.88
Latine	70.18	69.12
Black	0.00	4.41
Asian	1.75	1.47
Multiracial/Multiethnic	7.02	1.47
Unknown	12.28	17.65
Generational Status		
First-generation	19.30	33.82
Continuing generation	68.42	47.06
Unknown	12.28	19.12
College Major		
Biological Sciences	63.16	17.65
Other STEM discipline	22.81	45.59
Non-STEM	1.75	19.11
Unknown	12.28	17.65
Prior Research Experience		
No prior research experience	53.73	64.70
Prior research experience	24.56	17.65
Unknown	21.71	17.65

[a]The unknown category for each demographic feature is comprised of students who did not respond to the demographic survey.

and molecular biology CURE sections and three traditional laboratory course sections (n = 68) at a research intensive, Hispanic-Serving Institution in the Southwestern United States. These CURE and traditional laboratory course sections were offered during the Fall 2018 semester, were comprised largely of first-year students, and ran for the entire duration of the 15-wk semester. Student demography (Table 1) was similar across both CURE and traditional laboratory course sections based on gender identity and other self-reported characteristics – details of which were collected in a postsemester demographic survey. These statistics mirror the overall demographic makeup of the institution where the study was conducted, where Latine women constitute the majority of the student population.

Course Contexts

The CURE sections surveyed represent all biology CURE offerings at this university. More detailed information about these courses, including the meeting schedule; level of inquiry; and learning objectives, can be found in Table 1 of Esparza *et al.* (2020, p. 3). The CURE (C) courses included in our study met twice weekly in 3-h sessions. In these CUREs, students crafted testable hypotheses, designed experiments, and analyzed data to answer research questions within subdisciplines of cellular and molecular biology. Student research in these CUREs aligned with the research program of their respective faculty instructors, which included: zoonotic diseases (C1), evolutionary genetics (C2), and cancer biology (C3), the latter of which had parallel objectives and an identical course meeting structure as C1 and C2. The instructors of C2 and C3 randomly assigned students to fixed groups of four to five students, whom they worked alongside with for the entirety of the 15-wk semester. These fixed groups of students worked at the same lab benches throughout the term. There was little intentional group structuring done in C1 due to the layout of the classroom, and, as a result, students were free to move about the classroom and collaborate with peers during their research.

The three traditional laboratory sections involved in this study adopted a "cookbook" format common to introductory undergraduate biology laboratory curricula (Sundberg and Moncada, 1994). Students enrolled in the traditional laboratory courses were provided with discrete laboratory instructions and expected to complete one team-based exercise each session (e.g., loading an agarose gel, defining tonicity of solutions using brine shrimp eggs). The traditional labs met once weekly for approximately 2 h. Students in these three traditional laboratory sections worked together in randomly assigned and fixed groups of 4–5 students to complete the assigned laboratory exercises. These fixed groups of students sat at the same tables throughout the 15-wk semester.

The collection of human subjects data reported in this article was approved by The University of Texas at El Paso's Institutional Review Board under protocol #1117277. Written consent was provided voluntarily by all participants in the study.

Measures and Procedures

Name generator survey. To capture students' perceptions of their peers' performance in CUREs and traditional laboratory courses, we deployed an internally designed, open-ended name generator survey. In studies of social networks, name generator instruments are used to identify individuals with whom participants share a certain relationship or who have certain traits (Burt, 1984; Perry *et al.*, 2018). This survey allowed us to investigate student nomination structures (RQ1) and the influential factors (e.g., gender; RQ2) leading to nomination in CUREs and traditional laboratory courses. Specifically, this survey asked students to respond to the following prompts: 1) Please identify the student who, in your opinion, is the "star researcher" in this laboratory course; 2) Why did you select this student as your nomination?; and 3) Who, in this laboratory, would you say is most outspoken? Students were verbally informed that they could not select themselves or the graduate teaching assistant as their nominee before completing the survey.

To gain a representative account of student nomination behaviors, the name generator survey was given at weeks 4, 8, and 12 of each laboratory course. As indicated above, this survey asked students to nominate a single classmate whom they believed was most proficient in the course research (termed "star researcher"). Previous studies have adopted similar methods to examine students' estimations of their peers' performance in introductory biology lecture environments and mechanical engineering courses (Grunspan *et al.*, 2016; Salehi *et al.*, 2019b). Students were also asked to justify their selection via an open-ended name interpreter prompt. Name interpreters are often used in network studies to obtain more information about the individuals identified on the name generator instrument (Perry *et al.*, 2018). These qualitative justifications were then used to characterize the underlying reasons that students nominated their peers as proficient in CURE and traditional biology laboratory courses (RQ3).

We intentionally left the "star researcher" language ambiguous based on preliminary findings obtained from a version of the survey that we piloted in CURE and traditional laboratory courses in Fall 2017. Instead of asking students to list the name of a student who they believed was the "star researcher," this pilot survey asked students to rank their classmates on four dimensions: 1) technical skills, 2) experimental design skills, 3) content knowledge, and 4) leadership and collaboration. This was done on a scale of 1 to 5 for each of the dimensions, where 1 meant "low proficiency" and 5 meant "extreme proficiency." While each of the survey dimensions was defined for students on this survey, students mentioned that they found the survey confusing, and some ranked all of their classmates on the same level for each dimension (e.g., by ranking all classmates as a two on the four dimensions). Alternatively, students would rank one or a few students as a five on each dimension and neglect to fill out the remainder of the survey. Given these observations, we opted to allow students to qualitatively define their own reasons for viewing their peers to be proficient in the laboratory, rather than limiting students to the four dimensions that we had outlined. We also discovered through conversations with students in Fall 2017 that they viewed the investigations being conducted in their respective lab course – whether traditional or CURE – as constituting "research." In the traditional sections, this line of thought was reinforced by language contained in the course laboratory manual (Gonzalez, 2018). Thus, the term "star researcher" was seen as being most accessible and relatable to the students in our sample.

Students were asked to complete a pen and paper version of our name generator survey (see Supplemental Material S1) during class. Both students and instructors remained blind to the hypotheses being tested. During the first collection, students were told to take a sheet of cardstock, fold it in half, and write their names on both sides to create a "name tent." From the first collection onward, students were asked to display their "name tent" in such a way that it was easily visible to allow others to glance at their peers' names before nominating one of them as most proficient. Students were verbally instructed at each collection to write the full names of the persons whom they decided to nominate as proficient. When students submitted their name generator survey, author D.E. would examine the form to ensure that students provided full, legible names. Measures of student performance (i.e., exam scores, assignment grades) and course photo rosters were not openly available to the class at large, and students were only able to see their own grades via the university's Blackboard online learning management system.

Expanded Experimental Design Ability Tool. As both course types engaged students in experimentation, we implemented the Expanded Experimental Design Ability Tool (E-EDAT; Brownell *et al.*, 2014) to better understand the factors – aside from student gender identity – that informed students' receipt of a nomination (RQ2). While students in both the traditional laboratories and CUREs engaged in experimental design tasks, these tasks were of greater emphasis in the CUREs. These tasks included activities such as identifying research questions/hypotheses, discerning relevant variables, selecting appropriate research methods, and analyzing data, each of which is assessed by the E-EDAT. More specifically, the E-EDAT presents students with an open-ended prompt that asks them to design an experiment to test the efficacy of an herbal product, ginseng, in promoting endurance. Primary author D.E. scored responses to the E-EDAT using the rubric published alongside the instrument. To ensure the reliability and consistency of our scoring, corresponding author J.O. also analyzed student E-EDAT responses, and we subsequently calculated Cohen's kappa, a statistic used to assess the level of agreement between two raters while taking random chance of agreement into account (Cohen, 1960; McHugh, 2012). A high level of agreement ($\kappa = 0.914$, $p < 0.001$) was observed between the two raters. We incorporated students' E-EDAT scores into our models (see *Analysis* section below) to serve as a proxy for student performance. By doing this, we were able to determine the relationship between experimental design ability and the likelihood of being nominated as "star researcher" by others in the course. The E-EDAT was implemented in postsemester format in both the CURE and traditional laboratory courses.

Laboratory Course Assessment Survey. The laboratory course assessment survey (LCAS) is a 17-item measure of students' perceptions of the degree to which key design features (collaboration, discovery/broader relevance, iteration) were present in their biology laboratory course (Auchincloss *et al.*, 2014). The extent to which each of these constructs is integrated into a lab course is dependent on the level of inquiry of the lab, and previous work has established validity evidence verifying the use of the LCAS to distinguish between undergraduate student perceptions of these design features in CUREs and traditional laboratory environments (Corwin *et al.*, 2015). We implemented the LCAS at the end of the semester in the CURE and traditional laboratory courses to verify the degree to which students felt that collaboration was encouraged in their respective laboratory course.

Analysis

Network visualization. To address our first research question (RQ1), we created sociograms – visual representations of the relationships between people in a social network. Nodes are scaled by the number of nominations that a student received, known as in-degree. In-degree is a directed version of degree centrality and is one of the most commonly used measures to determine the influence or importance of a node within a network (Borgatti and Everett, 2006). Each of the sociograms presented in this article was constructed using the packages *ggraph* (Pedersen *et al.*, 2017) and *igraph* (Csardi and Nepusz, 2006) in the R statistical computing environment (v. 4.1.1; R Core Team, 2020).

Exponential random graph models to determine influential nomination factors. Similar to Salehi *et al.* (2019b) and Grunspan *et al.* (2016), we constructed exponential-family random graph models (ERGMs) to address the second research question (RQ2). Through the use of ERGMs, we were able to determine the factors (e.g., nominee gender) that influenced the nomination behaviors of students in each of the CURE and traditional laboratory contexts involved in this research. ERGMs are used to predict the formation of an edge (i.e., a nomination) between two nodes based on: 1) receiver effects, or the characteristics of the nodes receiving nominations (e.g., gender identity); 2) homophily effects, or the tendency for an individual to nominate a peer who shares similar characteristics (e.g., same lecture course) as themselves; and 3) structural effects, or the factors that influence the overall structure of a network (McPherson *et al.*, 2001; An, 2022). These effects are included as parameters in the ERGMs, which use Markov chain Monte Carlo maximum likelihood estimation (MCMC-MLE) to simulate a distribution for each of these parameters. Following, the mean of the simulated distribution is compared with the observed network data to understand whether the included parameters differ between the simulated and observed networks (Krivitsky, 2012).

Our initial ERGMs included: 1) an edge term, which controls for the density of the network and functions as the intercept for the ERGM (Morris *et al.*, 2008); and 2) a reciprocity term, which allows us to understand mutuality in nominations. However, we found evidence of sample statistic autocorrelation MCMC diagnostics in our CURE and traditional ERGMs as well as unbalanced density plots in our CURE model (Handcock *et al.*, 2003). Likewise, the goodness-of-fit for both models, while appropriate for the in-degree statistics, was poor when compared with other sample statistics (Hunter *et al.*, 2008a). These diagnostics, overall, suggested only moderate convergence of our models and potential degeneracy – a common issue when implementing ERGMs for the analysis of empirical networks (Snijders, 2002). In response, we included three geometrically weighted terms, which address issues of model degeneracy due to degree distributions and account for more complex dependencies in the edge formation process (Snijders *et al.*, 2006; Koskinen & Daraganova, 2013; Li and Carriere, 2013). These included a term for: 3) geometrically weighted edgewise-shared partners (*GWESP*), which accounts for the tendency for adjacent nodes to form triads (i.e., for students to nominate in a "triangle" or cluster); 4) geometrically weighted indegree (*gwidegree*); and 5) geometrically weighted outdegree (*gwodegree*). In addition to aiding model specification, the latter two of our structural terms are often used in research studies that model popularity effects (e.g., the tendency for a few individuals to receive a disproportionate number of nominations) and activity spread (the tendency for people to send a disproportionate number of nominations) of a social system and, thus, are fitting to answer our research questions (Cillessen *et al.*, 2011; Todd *et al.*, 2020). Further, receiver-effect parameters included gender identity and E-EDAT score. Homophily effect parameters included a term for lecture homophily, which allowed us to determine whether greater social contact and/or physical proximity between students who shared the same lecture section influenced student nomination patterns in the laboratory environment (Wellman, 1996). After incorporating these terms, our CURE and traditional models passed each of the MCMC diagnostic tests and exhibited improved goodness-of-fit as compared with our initial models (see Supplemental Material S2). We used the *ergm* package to estimate each of the ERGMs presented in this research (Hunter *et al.*, 2008b).

Qualitative analysis of open-ended responses. Primary author D.E. inductively coded participant responses to the name interpreter question to identify emergent themes in the data (Creswell, 2007) and, thus, to characterize the reasons why students nominated their peers as proficient. To ensure the reliability and consistency of our coding scheme, corresponding author J.O. coded 50% of the qualitative dataset, and we assessed interrater reliability by calculating Cohen's kappa (Cohen, 1960). Both coders achieved moderate-to-strong agreement (κ = 0.793, $p < 0.001$). Interrater agreement disaggregated across each code can be found in Supplemental Material S3.

RESULTS

RQ1. Nomination Structures Differ Between CURE and Traditional Laboratory Courses

We initially hypothesized that nominations would be more evenly distributed among students in CUREs given the higher degree of collaboration thought to be facilitated by CURE instruction (Auchincloss *et al.*, 2014). A Wilcoxon signed-rank test confirmed no significant difference ($p = 0.168$) between the average number of nominations received by CURE students ($M_{\text{CURE}} = 2.65$, $SD = 5.53$) and traditional students ($M_{\text{trad}} = 2.50$, $SD = 2.98$). However, there was greater variation in the number of nominations received among CURE students across CURE sections. Upon visual inspection of the sociograms representing student nomination behaviors (Figure 1), it appears this variation can be attributed to select students in CUREs who accrued substantially more nominations than their peers throughout the semester, thereby falsifying our hypothesis. In many cases, it further appears that several students allocated all three of their nominations to a single student, as is best observed in CURE 2 (C2) and CURE 3 (C3; Figure 1).

Notably, this effect changed over time on a course-by-course basis (see Supplemental Material S4). For instance, several students in C1 initially nominated a student of unknown gender at time point one, although they gradually began nominating others at time points two and three. Conversely, students in C2 and C3 initially had more variation in who they nominated at time point one but reached a consensus as a course on who they viewed as most proficient – a single "superstar" woman student in both cases – throughout time points two and three. Notably, in C3, men in the course largely nominated the other men at time point one. However, at time point three, none of the men nominated other men and, instead, nominated women in the course.

Still, there does appear to be variation in *who* students found to be most proficient in their laboratory sections across both CURE and traditional courses, with some students even nominating different peers at each time point. That said, it appears that nominations were generally evenly distributed, albeit clustered, across students in traditional labs. As such, it is difficult to determine who the "star researchers" were in these courses, wherein no one student emerged as the "[super]star researcher" (as was the case, for instance, in C2). Sociograms representing each course nomination network across each time point can be viewed in Supplemental Material S4.

RQ2. Gender does not Appear to Impact Nomination Patterns in CURE or Traditional Laboratory Courses

ERGM results for the aggregate CURE and traditional student nominations are shown in Table 2. The signs of the estimates (±) indicate the direction and magnitude of the effect of each of the parameters on the log-odds of a nomination (i.e., edge). For receiver (i.e., woman nominee, E-EDAT score) and homophily (i.e., lecture homophily) effects, a significant positive coefficient

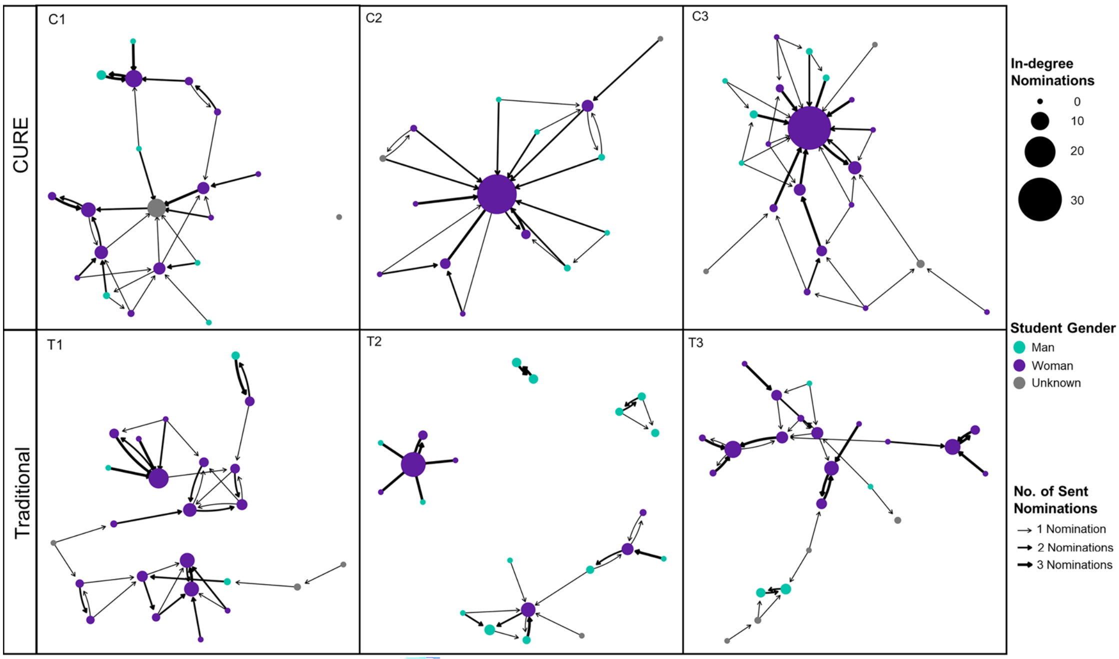

FIGURE 1. Sociograms for CURE and traditional courses. Nodes (circles) are sized by the number of nominations (in-degree) that students obtained throughout the semester. Edges (arrows) are sized by the number of nominations allocated by a single node, ranging from one nomination to three nominations throughout the semester. The edges are directional and have an arrow that signifies the flow of nominations (i.e., who sent the nomination to whom).

indicates an increased likelihood of the occurrence of a tie for a parameter, and a significant negative coefficient indicates a decreased likelihood of the occurrence of a tie for a parameter. Coefficient interpretations for structural parameters are more complex and are discussed alongside our results below.

Notably, the receiver effect for gender identity was not significant across any of the CURE or traditional laboratory course networks, indicating no influence of gender identity in our study context. These results suggest that, at least in the laboratory courses examined, students' gender identity does not influence whether they are perceived as apt in those biology laboratory environments, thereby falsifying our hypothesis that gender identity would relate to student perceptions of aptitude. Likewise, neither the receiver effect for E-EDAT score nor the homophily effect for lecture section significantly influenced student perceptions of aptitude in the CURE or traditional courses.

TABLE 2. Exponential Random Graph Models to understand nomination structures in CURE and traditional laboratory courses

Parameter	CURE 1 Est. (Std. Error)	CURE 2 Est. (Std. Error)	CURE 3 Est. (Std. Error)	Traditional 1 Est. (Std. Error)	Traditional 2 Est. (Std. Error)	Traditional 3 Est. (Std. Error)
Edges	**–2.76 (0.92)****	–0.98 (0.75)	**–3.38 (0.91)*****	**–7.27 (1.31)*****	**–6.50 (1.30)*****	**–5.38 (1.36)*****
Mutuality	**1.88 (0.76)***	1.30 (0.81)	–0.96 (1.25)	**1.96 (0.70)****	**3.82 (0.90)*****	**3.06 (0.70)*****
Woman Nominee	–0.04 (0.27)	0.88 (0.54)	0.20 (0.28)	0.63 (0.72)	0.24 (0.35)	0.97 (1.04)
Lecture Homophily	–0.17 (0.17)	–0.39 (0.60)	0.34 (0.48)	0.04 (0.38)	–0.54 (0.68)	–1.04 (0.72)
E–EDAT Score	–0.13 (0.09)	–0.09 (0.05)	0.02 (0.04)	0.04 (0.07)	0.13 (0.09)	–0.04 (0.05)
Nomination Popularity (*gwidegree*)	**–2.11 (0.74)****	**–3.26 (1.05)****	**–2.76 (0.64)*****	–0.10 (0.71)	–1.67 (0.87)	–1.11 (0.78)
Nomination Activity (*gwodegree*)	**3.96 (1.69)***	– [a]	**3.77 (1.62)***	**7.54 (2.72)****	**6.22 (1.99)****	**4.23 (1.50)****
Triangular Nominations (*GWESP*)	0.12 (0.41)	–0.09 (0.27)	0.67 (0.42)	**1.37 (0.33)*****	0.75 (0.39)	0.63 (0.34)

Significance Codes:*$p < 0.05$,**$p < 0.01$,***$p < 0.001$.
[a]The Nomination Activity term was removed in the ERGM for CURE 2 given high observed collinearity between Nomination Activity and the edges terms.

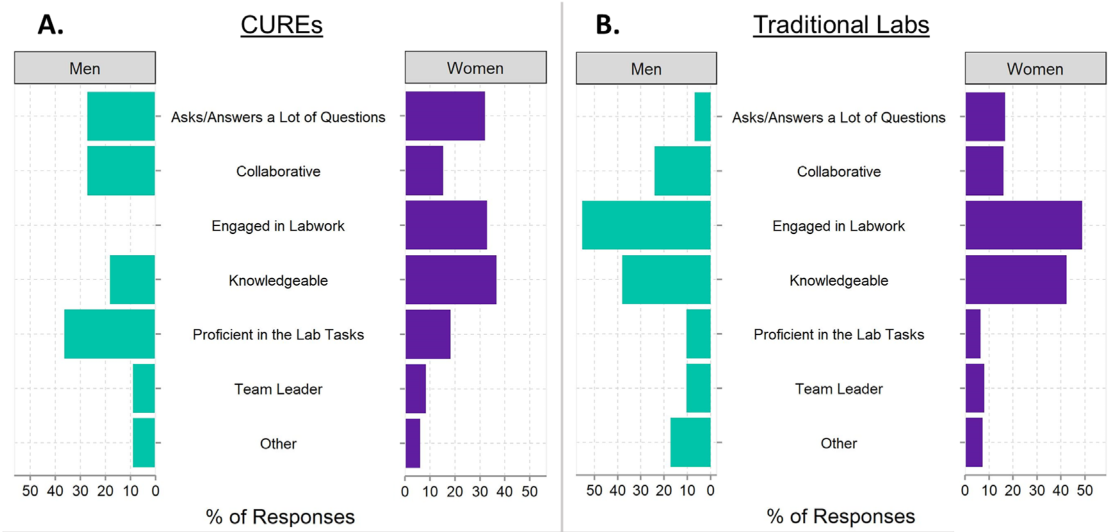

FIGURE 2. Summary of students' reasons for nominating peers in CURE and traditional laboratory courses stratified by gender identity. Codes were normalized based on the number of nominations that were allocated to men and women within each laboratory context, respectively. A full list of the codes and their full descriptions can be found in Supplemental Material S3.

While gender identity did not influence student nomination behaviors, we found that all five structural effects significantly impacted how students nominated their peers in both CURE and traditional courses. First, the edges estimate was significant and negative in two of the CURE and all traditional courses – a common result characteristic of real-life networks, suggesting that student nominations were not allocated randomly (Sigler *et al.*, 2021). The mutuality estimates were significant and positive in each of the traditional laboratory courses. This result suggests that, in traditional courses, students were likely to reciprocate nominations that they received from their peers. This is especially evident when examining the traditional course sociograms (Figure 1), wherein students often nominated peers who nominated them – lending further credence to this result. This was less often present in the CURE networks, wherein students nominated one or a few students throughout the semester as "star researcher." We found evidence for triangular nomination structures in a single traditional course, suggesting that nominations in this course tended to flow between small clusters of students.

We observed a prominent popularity effect in each of the CURE courses involved in this research. This is made evident by the negative and significant "Nomination Popularity" estimate – a result suggesting that the CURE nomination networks are centralized around a few students who accrued a disproportionate number of nominations throughout the semester (Levy, 2016). The topography of the CURE sociograms (Figure 1) supports these results. For instance, a few students seem to have been recognized by most of their peers as apt – becoming "global star researchers" in the course. This is most apparent in the networks for CURE 2 (C2) and CURE 3 (C3). The effect of nomination activity was positive and significant for all examined traditional laboratory and CURE courses. These coefficients indicate that the outgoing nominations are decentralized and, overall, homogenous – in other words, all students sent similar numbers of nominations throughout the semester (Levy, 2016; Kaven *et al.*, 2021). This result is unsurprising given that all students had an equal number of nominations to send to their peers – one per time point for a total of three across the entire semester.

RQ3. A Variety of Factors Influenced Student Nominations in CURE and Traditional Courses

We collected information on students' reasons for nominating their peers in CURE and traditional courses. An overall summary of these patterns – normalized by the course type and student gender – is presented in Figure 2.

Women and Men were Nominated for Different Reasons in CUREs. From a descriptive standpoint, nomination justifications differed somewhat by gender in CURE courses (Figure 2A). In CUREs, women were most often recognized for their knowledge (36.64%) of the course content – that is, their demonstration of an understanding of the laboratory content or ability to accrue content knowledge quickly. Pertinently, while women (M_{women} = 5.18, SD = 2.92) in CUREs performed better on the E-EDAT than men (M_{men} = 3.63, SD = 2.13), this difference was not significant (p = 0.071). In describing their peer's knowledge, one CURE student said:

> "*She seems to grasp concepts pretty easily* and apply them to the whole purpose of why we do things."

This student's justification suggests that the nominee was able to quickly learn and mobilize her understanding of laboratory content. In contrast, men in CUREs were most often nominated based on their appearance of proficiency in laboratory tasks – such as by completing laboratory tasks quickly, operating the experimental apparatuses well, or being methodical during the research process. When nominating a man peer, one CURE student said:

"He is methodical in his procedures. He is a bit of a micromanager, which does well to ensure readings are likely to be accurate…"

Both women (32.06%) and men (27.27%) in CUREs were nominated for their openness in asking and answering questions during laboratory sessions. For example, students justified their nominations by stating the following about their peers:

"She asks the most questions and is not afraid to ask things everyone else doesn't want to ask."

"[Student] is always asking questions if he is unsure of a procedure…"

As illustrated by the above quotes, another common theme in student nomination justifications was collaboration. Both men (27.27%) and women (15.26%) in CUREs were nominated for being collaborative. When nominated, collaborative "star researchers" were described in the following ways:

"He asks thorough questions and guides me and other classmates through experiments."

"She doesn't just listen when the teacher talks, she tries to break it down and fully comprehend. *She also loves helping others when ahead.*"

Nomination Patterns were Mostly Uniform in Traditional Courses. While student justifications for their nominations slightly differed by gender in CUREs, traditional students' reasons for allocating a nomination were, descriptively speaking, relatively uniform across codes when stratified by gender (Figure 2B). Women (48.91%) and men (55.17%) in traditional courses were most often nominated by their peers for being engaged in the lab work – in other words, for expressing interest in the lab content; being "on task;" or preparing for activities before/during class. Exemplar quotes for this code are as follows:

"He is reliable and is detail-oriented. *Every lab he arrives prepared…*"

"*Because she seems to be so interested in every experiment,* she tries to understand every part of it and *she seems to enjoy lab experiments.*"

Like students in CUREs, women (16.05%) and men (24.14%) in traditional courses were described by their peers as collaborative, as evidenced by the following testimony:

"*Because he's a very helpful lab partner*, sometimes when I can't figure something out, he helps me out and usually knows the answers to my questions."

Women (42.34%) and men (37.93%) in traditional courses were also described as knowledgeable. Pertinently, E-EDAT scores did not significantly differ between men ($M_{men} = 4.00$, SD = 2.94) and women ($M_{women} = 3.78$, SD = 2.56) enrolled in traditional laboratory courses ($p = 0.803$). One student provided the following description of a peer:

"She is always informed about the subject at hand. She goes above what is asked in this class. *Knows more content than what is taught in class.*"

Notably, we observed that some students in traditional courses would nominate a peer solely because they were amongst the one or few people that they had talked to in their laboratory course. For instance, traditional students would describe nominating a classmate because they were "lab partners" or because they were "in their lab group." This phenomenon is represented by the following quote:

"She is in my group; I haven't had the chance to really interact with anyone outside my group. However, when we are working in the lab, she is always very helpful."

This latter observation suggests that traditional students may have had fewer opportunities to engage in collaborative practice, especially with people outside of their laboratory groups, than what was reported for CURE students. To explore this possibility, we performed Wilcoxon signed-rank tests to compare CURE and traditional students' responses on the LCAS. We found that students in the CUREs perceived a significantly higher degree of discovery/broader relevance ($M_{CURE} = 17.7$, SD = 5.0; $M_{trad} = 14.4$, SD = 5.1) and iteration ($M_{CURE} = 23.4$, SD = 5.2; $M_{trad} = 18.9$, SD = 5.9) as compared with students in traditional courses ($p < 0.001$ for both comparisons). Most pertinently, we found that the students in traditional courses perceived a significantly lower ($p < 0.01$) degree of collaboration ($M_{trad} = 21.00$, SD = 6.00) relative to students in CURE courses ($M_{CURE} = 24.30$, SD = 4.79).

DISCUSSION

Recent studies have discovered gender-based inequities in the undergraduate biology curriculum. For instance, women perform less well on biology exams and exhibit greater test anxiety as compared with men (Ballen *et al.*, 2017; Cotner *et al.*, 2020). Despite the growing literature on the role of gender identity in undergraduate biology learning environments, few studies have investigated the underlying mechanisms that result in such inequities. As such, we examined the following research questions: 1) What do student nomination structures look like in CUREs and traditional laboratory courses?; 2) Does a student's gender identity influence the probability that they will get nominated as proficient by their peers in CURE and traditional courses?; and 3) For what reasons do students view their peers as proficient, and are there differences in the way students justify their nominations based on the gender identity of the nominee?

RQ1. The Influence of Laboratory Course Structure on Students' Perceptions of their Peers

Prior work suggests that the broader contextual features of an environment can influence *how* individuals evaluate their peers and collaborators (Tziner *et al.*, 2008). In line with these findings, we conjecture that the structure of biology laboratory courses may influence how students view the aptitude of their peers. CUREs, by definition, engage students in research in a course setting (Auchincloss *et al.*, 2014). Like other types of higher-inquiry laboratory experiences, CUREs often forego conventional characteristics of traditional “cookbook” labs (e.g., confirmatory experiments) to engage students in laboratory-based curricula that better reflect the authentic epistemology of science (Chinn and Malhotra, 2002). Recent studies have suggested that laboratory courses that adopt higher forms of inquiry – such as CUREs – encourage collaboration and diversified group roles by necessitating the division of labor to complete lab assignments (Quinn *et al.*, 2020). Collaboration is a key design element of CUREs, and previous studies reliably indicate that students perceive greater collaboration in CUREs than in traditional laboratory courses (Corwin *et al.*, 2015; Esparza *et al.*, 2020).

In the present study, we found similar results regarding students’ perceptions of collaboration, wherein students in CUREs perceived a significantly higher degree of collaboration. We posit that the greater degree of collaboration facilitated by the CURE courses provided more opportunities for students to gain an informed perspective of the strengths of their peers. Management research indicates similar findings, wherein raters with sufficient information about an individual’s abilities are better suited to make informed decisions about their aptitude (Tosi and Einbender, 1985). Thus, as students work with their peers during their course research, it is likely that they can form a more developed idea of who is proficient in the course activities.

Notably, we also observed changes in CURE student nomination patterns over time, wherein students initially may have viewed certain peers as more proficient early in the semester but changed who they nominated as the course progressed (Figure 1; see Supplemental Material S4). Therefore, it is likely that student perceptions of their peers’ aptitude may also change over time as students see their peers operate within open-inquiry laboratory environments.

RQ2. Gender Identity does not Appear to Influence Student Nomination Patterns in the CUREs or Traditional Laboratory Courses Involved in this Study

In comparison to the research of Grunspan *et al.* (2016) – which found that men underestimate the knowledge of women in introductory biology lecture courses – we found that gender identity did not influence the log-odds of a student receiving a nomination in CURE or traditional laboratory courses. Likewise, performance on the E-EDAT did not influence the odds of a student’s receipt of a nomination and did not differ by student gender identity in CURE and traditional laboratory courses. Our results most directly align with those of Salehi *et al.* (2019b), who found no evidence of a gender bias in student perceptions of their peers across both a traditional and interactive offering of a mechanical engineering course. In our context, it does raise the question: why might we observe gender biases in student perceptions in biology lecture courses as compared with laboratory courses?

One potential explanation for our findings lies in the active and experiential nature of laboratory-based education. In laboratory courses – both traditional and discovery-based – students can, at least in part, perform and learn the practices of a scientist while engaging with their peers (Brownell *et al.*, 2012). Within our context, peer-to-peer interactions could have enabled students to make more informed decisions about their peers’ aptitude in the CURE and traditional laboratory courses than they would have been able to make in a lecture course, especially given that these laboratory sections had smaller enrollment than the corresponding lecture courses. Pertinently, recent research on STEM lecture courses indicates that smaller course sizes can promote gender equity in class participation (Ballen *et al.*, 2019). As such, it is possible that the smaller class sizes of our CURE and traditional laboratory courses contributed to more equitable class participation and, by extension, more informed and unbiased peer perceptions. Lastly, previous research in biology lecture courses shows that women will participate more often when there are a higher percentage of women students in attendance (Bailey *et al.*, 2020). It is therefore possible that the high percentage of women students in our CURE and traditional laboratory courses may have resulted in more equitable participation. Future research should aim to disentangle gender dynamics in peer perceptions of biology on a larger scale, perhaps including an examination of students’ level of comfort in participating across different biology teaching contexts (e.g., lecture, lab). Still, these results are encouraging and add further credence to arguments that CUREs can advance inclusivity in biology (Bangera and Brownell, 2014).

The structural effects of the CURE and traditional laboratory course social networks explained how nominations were allocated in these contexts. For example, we observed a prominent popularity effect in CUREs, as indicated by our sociograms (Figure 1) and the negative and significant geometrically weighted indegree coefficient in our ERGMs (Table 2). We posit that situational factors (e.g., time spent together, degree of collaboration, the potential for supportive vs. authoritative learning environments) may be contributing to such outcomes (Auchincloss *et al.*, 2014; Esparza *et al.*, 2020).

RQ3. Deconstructing the Reasons Students Nominated Peers in CURE and Traditional Laboratory Courses

In coding the justifications that students provided when nominating a peer as most proficient, we discovered that the reasons for nominations stayed relatively consistent in traditional courses. Interestingly, while we found no impact of gender identity on the log-odds of students receiving a nomination, the primary reasons students were nominated in CUREs were somewhat different based on gender identity. We conjecture that the structural features of these courses mediated variation in how nominations were justified amongst the women and men enrolled in CUREs and traditional laboratory courses. Specifically, we couch these arguments in constructionist perspectives of gender, which consider gender as a self-constructed aspect of an individual that is flexible, actively performed, and that may differ based on context and situation rather than being concrete and deterministic (West and Zimmerman, 1987).

Notably, we found no differences in students’ performance on the E-EDAT in CURE or traditional laboratory courses as a function of gender. Further, students’ performance on the E-EDAT had no impact on the log-odds of receiving a nomination. However, our qualitative results suggest that a greater proportion of nominations directed toward men in CUREs focused on their proficiency in lab tasks (e.g., operating experimental apparatuses) than was the case for women. In contrast, women in CUREs were, proportionately, most often recognized for their knowledge of biology, content taught in the lab, and the broader field of study emphasized in their CURE. Given these gender differences in the reasons students were most often nominated in CURE environments, it is plausible that the less-structured, open form of scientific inquiry supported in CUREs may influence how students divide and perform tasks. As opposed to traditional "cookbook" laboratory courses – which ask students to follow step-by-step instructions for confirmatory experiments – CUREs provide students with an opportunity to engage in more open and unguided forms of inquiry that allow for them to determine how best to address their research goals (Nadelson *et al.*, 2010; Brownell and Kloser, 2015). Recent studies have discovered that gender-based inequities in the division of laboratory tasks are more prominent in laboratory curricula that support less structured forms of inquiry – such as that which is facilitated in CUREs (Brownell and Kloser, 2015; Quinn *et al.*, 2020). These inequities are likely the result of the conflation of "doing science" with "doing gender," wherein students adopt stereotypical masculine or feminine roles when doing laboratory tasks (Danielsson and Linder, 2009; Holmes *et al.*, 2014; Gonsalves *et al.*, 2016). For instance, Quinn *et al.* (2020) found that women used laptop computers more often than men, while men used experimental apparatuses more often than women in inquiry-based laboratory environments. Similarly, Doucette *et al.* (2020) found that men often fell into "tinkerer" roles, wherein they did most of the experimental work in undergraduate physics laboratory courses. While these roles are necessary for the completion of the laboratory work, some have conjectured that who falls into these roles is the result of "gendering science," a process in which the "ideas, practices, epistemic values, and intuitions of science become masculinized and feminized," impacting who is ultimately viewed as an effective and apt scientist (Rutherford, 2020).

Relatedly, Quinn *et al.* (2020) found no indicators of gendered task division in traditional laboratory environments, a finding attributed to the increased pedagogical structure of traditional laboratory courses. Our qualitative results align with this phenomenon, wherein the percentage of responses across all codes for men and women was approximately equivalent when normalized by the overall number of nominations by student gender. Likewise, previous research on introductory biology lecture courses suggests that moderate course structure can improve the performance and engagement of all students (Eddy and Hogan, 2014). Thus, it may be helpful for instructors who teach biology laboratory courses to add or reinforce pedagogical structure through cooperative grouping (Heller *et al.*, 1992), in which students are assigned roles (e.g., principal investigator, data analyst, notebook manager, experimenters) that alternate each lab session. Similar methods have previously been deployed in CUREs and CURE-like spaces (e.g., Luckie *et al.*, 2013; Olimpo *et al.*, 2016). Further, instructors of CUREs and traditional labs can create equivalent peer groups based on gender identity (Lou *et al.*, 2000), which have been shown to support student performance and potentially curb gender inequities (Day *et al.*, 2016).

When normalized by the overall number of nominations allocated based on gender, we found that CURE students nominated women and men roughly equally for their propensity to ask and answer questions during the lab. The nonsignificant term for gender also supports the proportion of nominations focused on asking/answering questions being approximately equal across the gender binary seen in our research. Prior studies have illustrated that courses that promote active learning can help close demographic opportunity gaps in collegiate STEM learning environments (Theobald *et al.*, 2020; Freeman *et al.*, 2014). Yet, observational research indicates that women participate less than men in both didactic biology lecture courses (Eddy *et al.*, 2014) and biology lecture courses that implement active learning strategies (Aguillon *et al.*, 2020). Our findings provide insight into gendered participation in laboratory courses, suggesting more equitable gendered social participation in biology laboratory environments. These findings provide some empirical support for the hypothesis that biology laboratory courses, including CUREs, can help foster a more inclusive STEM landscape (Bangera and Brownell, 2014). Still, our findings require some cautious optimism; our qualitative results suggest that a small subset of students were nominated based, at least in part, on their ability to ask and answer questions. Given the Nomination Popularity effect observed in the CURE courses involved in this research, it may be fruitful for CURE instructors to employ inclusive teaching strategies such as waiting for a few students to volunteer responses to instructor-posed questions (Shahrill, 2013) and randomly calling on students (Eddy *et al.*, 2014; Knight *et al.*, 2016) to encourage the participation of a broader array of students.

LIMITATIONS AND FUTURE DIRECTIONS

A notable limitation of this work is that gender identity is treated as a binary despite the gender diversity observed across biology learning contexts (Cooper and Brownell, 2017; Haverkamp, 2021) and in STEM courses, broadly (Maloy *et al.*, 2022). Gender identity is a socially defined construct that refers to a person's inherent and deep identification with characteristics within the gender binary (e.g., man, woman) and/or gender expansive (e.g., nonbinary, gender fluid) identities (American Psychological Association, 2015). An individual's gender identity is performed and constructed gradually throughout their lives and can differ from the sex (e.g., male, female) that they are assigned at birth (Polderman *et al.*, 2018). Our research does not capture the holistic complexity of gender as a social construct in biology laboratory environments. This is because students in the present study solely identified within the gender binary despite being given the option on the demographic survey to identify as nonbinary. Additionally, the completeness of our data on student gender identity may also impact the results of this research. While approximately 90% of students reported their gender identity (CURE percentage = 89.5%, traditional percentage = 88.2%), some students either dropped the course or were not present when the postsemester survey was administered, which contained the E-EDAT, LCAS, and demographic survey. ERGMs require that the nodal attribute data (i.e., data on the characteristics of each student) is complete. While we are able to categorize these students as "unknown gender" for the sociograms (Figure 1), we did not have data on these students' E-EDAT scores and lecture sections – both parameters used in our ERGMs. As a result, students categorized as "unknown gender" were removed from the ERGM analyses. Future research should be done to explore the experiences of cisgender and gender-expansive students in biology laboratory contexts to support the development of more inclusive laboratory education.

It is imperative to acknowledge that gender identity is only one variable impacting how students present, perform, act, and interact within college STEM contexts. Several previous studies indicate that racial identity and student major can influence the social relationships that students form during college (Mayer and Puller, 2008; Wimmer and Lewis, 2010; Xu *et al.*, 2019). Our study was conducted at a Hispanic-Serving Institution, wherein 70% or more of our sample identified with a marginalized race or ethnicity – a percentage not reflective of the population of students enrolled in U.S. biology undergraduate degree programs (National Science Foundation, 2017). As such, this lack of variation in students' racial and ethnic identities may influence our analyses. Therefore, future research should investigate students' perceptions of their peers' aptitude in more diverse biology laboratory contexts. In addition, prior research has revealed that attrition from STEM degree programs is higher for students who identify with marginalized races and/or ethnicities as compared with White students (Riegle-Crumb *et al.*, 2019). The impact of inequities and stereotyping can compound when students identify with two or more marginalized identities (Hazari *et al.*, 2013; Saw *et al.*, 2018). Accordingly, further studies of students' evaluations of their peers' aptitude or knowledge should take an intersectional approach to determine whether certain students are underestimated by their peers based on their racial/ethnic *and* gender identity in college STEM lecture and laboratory environments.

Another potential limitation is the use of the E-EDAT as a proxy for student performance. Salehi *et al.* (2019b) and Grunspan *et al.* (2016) used student grades to understand the impact of student performance in their respective lecture courses on student perceptions of their ability. In our case, the measures used to determine student grades across the various sections of both the CURE and traditional laboratory courses were highly variable. For instance, each of the CURE sections was predominantly graded based on students' attendance and the effort they placed into their course research, resulting in most of these students receiving an "A" in these courses. On the other hand, the traditional laboratory course grades were calculated based on students' performance on quizzes, assignments, laboratory reports, presentations, attendance, and the maintenance of a laboratory notebook. Given this variation, we found grades to be an unreliable means to evaluate student performance objectively and consistently across CURE and traditional laboratory courses. As a result, we implemented the E-EDAT, although we recognize that this is a potentially imperfect method of holistically measuring student performance in CURE and traditional laboratory courses.

Prior research has established that sociometric nominations are highly valid and reliable approaches to understanding the popularity of individuals in a social system (Cillessen *et al.*, 2011). While social network analysis is a powerful tool to elicit social patterns within learning contexts, issues such as students forgetting interactions and/or the names of their peers – known as recall bias – are common (Brewer, 2000; Wright and Pescosolido, 2002; Rice *et al.*, 2014). To minimize the risk of recall bias in soliciting nominations, students were instructed to create "name tents" (see *Materials and Methods: Measures & Procedures*) and asked to display them in such a way that allowed their peers to view each other's names before submitting their nominations. This process was done due to institutional review board restrictions regarding the distribution of course rosters. Still, it is wholly possible that: 1) students may not have been able to see the name tent given the geography of the laboratory, and/or 2) students may not have been able to recall the name of a classmate that they *would have* nominated as proficient or outspoken if they were absent from class that day.

One further potential limitation of this work is its scope, specifically the sample size and the number of laboratory contexts studied. Defining sample size in social networks is a subject of debate, and several factors are crucial when determining effective sample size (Krivitsky and Kolaczyk, 2015). Notably, while issues of degeneracy and instability should be considered when using ERGMs, the models used in this research all fit the data (see Supplemental Material S2), and each converged. Like other statistical methods, the small sample size in our research is likely reflected in the p values and standard errors of our findings. Yet, it is important to note that prior studies employing ERGMs have been conducted on networks of a similar size to those presented in this study. For example, one study employed ERGMs on a single physics course containing 19 students (Wells, 2019). Further, several studies have explored the efficacy of different approaches when estimating ERGMs using the Lazega network, a network describing collaborations within a law firm containing 36 individuals (Hunter and Handcock, 2006; Van Duijn *et al.*, 2009; Box-Steffensmeier *et al.*, 2018). While methods exist to implement ERGMs for small networks, such as the *ergmito* package (Vega Yon *et al.*, 2021), we found that our networks contained too many unique dyads or were too large to fit ERGMs using this method. Thus, we contend that the issue is less one of sample size and more one associated with the impact of sample size on the validity of the findings.

Finally, it is critical to note that the frequency at which students were nominated based on certain traits may not necessarily reflect the actual behaviors of students in undergraduate biology laboratory courses. A student's nomination is rooted in their perception of a peer within the context of the undergraduate laboratory course, and, thus, may or may not accurately reflect the behaviors of the nominee. In CUREs, we found that men were more often recognized for their experimental and technical skills while women were lauded for their knowledge of the course content when justifications were normalized by course type and gender. This effect was not apparent in traditional courses as, when normalized, the reasons for nominations were approximately equally distributed across women and men who were nominated. Given these findings and prior research on the influence of laboratory structure on student participation and task division (Neill *et al.*, 2019; Quinn *et al.*, 2020), future research should investigate the division of tasks in CUREs and traditional laboratories. Such a study could implement existing video observation protocols (e.g., Day *et al.*, 2016) to characterize the types of tasks that are traditionally "gendered" (e.g., equipment use, note-taking) in undergraduate biology laboratory courses.

ACKNOWLEDGMENTS

We are indebted to each of the individuals who participated in this study; their responsiveness and willingness to participate ultimately aided in our understanding of gender equity in undergraduate biology labs. We thank Lillian Grace Sumption Senn for her thoughtful feedback on early drafts of the manuscript. We thank Amy E. Wagler for her advice on statistical analyses.

REFERENCES

Aguillon, S. M., Siegmund, G. F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. *CBE—Life Sciences Education*, *19*(2), ar12.

American Psychological Association (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American psychologist*, *70*(9), 832–864.

An, W. (2022). Friendship network formation in Chinese middle schools: patterns of inequality and homophily. *Social Networks*, *68*, 218–228.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., …& Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, *13*(1), 29–40.

Bailey, E. G., Greenall, R. F., Baek, D. M., Morris, C., Nelson, N., Quirante, T. M., … & Williams, K. R. (2020). Female in-class participation and performance increase with more female peers and/or a female instructor in life sciences courses. *CBE—Life Sciences Education*, *19*(3), ar30.

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS One*, *12*(10), e0186419.

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.

Bangera, G., & Brownell, S. E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sciences Education*, *13*(4), 602–606.

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.

Bernard, R. E., & Cooperdock, E. H. (2018). No progress on diversity in 40 years. *Nature Geoscience*, *11*(5), 292–295.

Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466–484.

Box-Steffensmeier, J. M., Christenson, D. P., & Morgan, J. W. (2018). Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. *Political Analysis*, *26*(1), 3–19.

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, *22*(1), 29–43.

Brownell, S. E., & Kloser, M. J. (2015). Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology. *Studies in Higher Education*, *40*(3), 525–544.

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. (2012). Undergraduate biology lab courses: Comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. *Journal of College Science Teaching,*, *41*(4), 36–45.

Brownell, S. E., Wenderoth, M. P., Theobald, R., Okoroafor, N., Koval, M., Freeman, S., … & Crowe, A. J. (2014). How students think about experimental design: Novel conceptions revealed by in-class activities. *BioScience*, *64*(2), 125–137.

Burt, R. S. (1984). Network items and the general social survey. *Social Networks*, *6*(4), 293–339.

Chen, X. (2013). *STEM Attrition: College Students' Paths into and out of STEM Fields*. *Statistical Analysis Report. NCES 2014-001*. Washington, DC: National Center for Education Statistics. Retrieved October 6, 2023, from www.nces.ed.gov/pubs2014/2014001.rev.pdf

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

Cillessen, A. H., Schwartz, D., & Mayeux, L.((2011). *Popularity in the peer system*. Guilford Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cooper, K. M., & Brownell, S. E. (2017). Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA students. *CBE—Life Sciences Education*, *15*(3), ar37.

Corwin, L. A., Runyon, C., Robinson, A., & Dolan, E. L. (2015). The laboratory course assessment survey: A tool to measure three dimensions of research-course design. *CBE—Life Sciences Education*, *14*(4), ar37.

Cotner, S., Jeno, L. M., Walker, J. D., Jørgensen, C., & Vandvik, V. (2020). Gender gaps in the performance of Norwegian biology students: The roles of test anxiety and science confidence. *International Journal of STEM Education*, *7*(1), 1–10.

Creswell, J. W. (Ed.) (2007). Data analysis and interpretation. In Creswell, J. W., (Eds.), *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed., pp. 147–176).Thousand Oaks, CA: Sage Publications.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Danielsson, A. T., & Linder, C. (2009). Learning in physics by doing laboratory work: Towards a new conceptual framework. *Gender and Education*, *21*(2), 129–144.

Day, J., Stang, J. B., Holmes, N. G., Kumar, D., & Bonn, D. A. (2016). Gender gaps and gendered action in a first-year physics laboratory. *Physical Review Physics Education Research*, *12*(2), 020104.

Deaux, K. (1984). From individual differences to social categories: Analysis of a decade's research on gender. *American Psychologist*, *39*(2), 105.

Doucette, D., Clark, R., & Singh, C. (2020). Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning. *European Journal of Physics*, *41*(3), 035702.

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), 020106.

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.

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.

Esparza, D., Wagler, A. E., & Olimpo, J. T. (2020). Characterization of instructor and student behaviors in CURE and non-CURE learning environments: Impacts on student motivation, science identity development, and perceptions of the laboratory experience. *CBE—Life Sciences Education*, *19*(1), ar10.

Finley, A., & McNair, T. (2014). Assessing underserved students' engagement in high-impact practices. Retrieved March 11, 2022, from https://vtechworks.lib.vt.edu/bitstream/handle/10919/87004/AssessingUnderserving Students.pdf?sesequen=1

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*, *111*(23), 8410–8415.

Ganley, C. M., George, C. E., Cimpian, J. R., & Makowski, M. B. (2018). Gender equity in college majors: Looking beyond the STEM/Non-STEM dichotomy for answers regarding female participation. *American Educational Research Journal*, *55*(3), 453–487.

Gonsalves, A. J., Danielsson, A., & Pettersson, H. (2016). Masculinities and experimental practices in physics: The view from three case studies. *Physical Review Physics Education Research*, *12*(2), 020120.

Gonzalez, H. (2018). *Topics in the Study of Life Laboratory Manual*. Plymouth, MI: Hayden-McNeil.

Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS One*, *11*(2), e0148405.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). ERGM: Fit, simulate and diagnose exponential-family models for networks, Version 2.1. URL: Retrieved from http://statnetproject.org

Haverkamp, A. E. (2021). Transgender and gender nonconforming undergraduate engineering students: Perspectives, resiliency, and suggestions for improving engineering education.

Hazari, Z., Sadler, P. M., & Sonnert, G. (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. *Journal of College Science Teaching*, *42*(5), 82–91.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627–636.

Holmes, N. G., Roll, I., & Bonn, D. A. (2014). Participating in the physics lab: Does gender matter? *Physics in Canada*, *70*(2), 84–86.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), nihpa54860.

Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of computational and graphical statistics*, *15*(3), 565–583.

Kaven, E., Kaven, I., Gómez-Zará, D., DeChurch, L., & Contractor, N. (2021). Assessing how team task influences team assembly through network analysis. In Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., & Sales-Pardo, M. (Eds.), *Complex Networks & Their Applications IX. COMPLEX NETWORKS. Studies in Computational Intelligence* (Vol. 944). Cham, CH: Springer.

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), ar56.

Kokott, K., Lengersdorf, D., & Schlüter, K. (2018). Gender construction in experiment-based biology lessons. *Education Sciences*, *8*(3), 115.

Koskinen, J., & Daraganova, G. (2013). Exponential random graph model fundamentals. In Lusher, D., Koskinen, J., & Robins, G. (Eds.), *Exponential random graph models for social networks: Theory, methods, and applications* (Vol. 35, pp. 49–76). New York, NY: Cambridge University Press.

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, *6*, 1100.

Krivitsky, P. N., & Kolaczyk, E. D. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science*, *30*(2), 184.

Levy, M. A. (2016). gwdegree: Improving interpretation of geometrically-weighted degree estimates in exponential random graph models. *Journal of Open Source Software*, *1*(3), 36.

Li, Y., & Carriere, K. C. (2013). Assessing goodness of fit of exponential random graph models. *International Journal of Statistics and Probability*, *2*(4), 64.

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *The Journal of Educational Research*, *94*(2), 101–112.

Luckie, D. B., Smith, J. J., Cheruvelil, K. S., Fata-Hartley, C., Murphy, C. A., & Urquhart, G. R. (2013). The "Anti-cookbook laboratory": Converting "canned" introductory biology laboratories to multi-week independent investigations. *Proceedings of the Association for Biology Laboratory Education*, *34*, 196–213.

Maloy, J., Kwapisz, M. B., & Hughes, B. E. (2022). Factors influencing retention of transgender and gender nonconforming students in undergraduate STEM majors. *CBE—Life Sciences Education*, *21*(1), ar13.

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., ... & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open*, *3*(4), 2332858417743754.

Mayer, A., & Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of public economics*, *92*(1-2), 329–347.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software*, *24*(4), 1548.

Nadelson, L. S., Walters, L., & Waterman, J. (2010). Course-integrated undergraduate research experiences structured at different levels of inquiry. *Journal of STEM Education: Innovations and Research*, *11*(1), 27–44.

National Science Foundation, National Center for Science and Engineering Statistics (2017). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017.* Special Report NSF 17-310, (p. 37). Arlington, VA: National Science Foundation. www.nsf.gov/statistics/wmpd/

Neill, C., Cotner, S., Driessen, M., & Ballen, C. J. (2019). Structured learning environments are required to promote equitable participation. *Chemistry Education Research and Practice*, *20*(1), 197–203.

Olimpo, J. T., Fisher, G. R., & DeChenne-Peters, S. E. (2016). Development and evaluation of the Tigriopus course-based undergraduate research experience: Impacts on students' content knowledge, attitudes, and motivation in a majors introductory biology course. *CBE—Life Sciences Education*, *15*(4), ar72.

Pedersen, T. L., Pedersen, M. T. L., LazyData, T. R. U. E., Rcpp, I., & Rcpp, L. (2017). *Package 'ggraph'. Retrieved on March 12*, 2022.

Perry, B. L., Pescosolido, B. A., & Borgatti, S. P. (2018). *Egocentric network analysis: Foundations, methods, and models* (pp. 68–108). New York, NY: Cambridge University Press.

Polderman, T. J., Kreukels, B. P., Irwig, M. S., Beach, L., Chan, Y. M., Derks, E. M., ... & Davis, L. K. (2018). The biological contributions to gender identity and gender diversity: Bringing data to the table. *Behavior genetics*, *48*(2), 95–108.

Puaca, L. M. (2013). *Searching for scientific womanpower: Technocratic feminism and the politics of national security, 1940-1980* (pp. 9–42). Chapel Hill, NC: UNC Press Books.

Quinn, K. N., Kelley, M. M., McGill, K. L., Smith, E. M., Whipps, Z., & Holmes, N. G. (2020). Group roles in unstructured labs show inequitable gender divide. *Physical Review Physics Education Research*, *16*(1), 010129.

R Core Team (2020). R: A language and environment for statistical computing.

Rice, E., Holloway, I. W., Barman-Adhikari, A., Fuentes, D., Brown, C. H., & Palinkas, L. A. (2014). A mixed methods approach to network data collection. *Field Methods*, *26*(3), 252–268.

Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher*, *48*(3), 133–144.

Rutherford, A. (2020). Doing science, doing gender: Using history in the present. *Journal of Theoretical and Philosophical Psychology*, *40*(1), 21.

Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., ... & Ballen, C. J. (2019a). Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. *Frontiers in Education*, *27*(4), 107.

Salehi, S., Holmes, N. G., & Wieman, C. (2019b). Exploring bias in mechanical engineering students' perceptions of classmates. *PLoS One*, *14*(3), e0212477.

Saw, G., Chang, C. N., & Chan, H. Y. (2018). Cross-sectional and longitudinal disparities in STEM career aspirations at the intersection of gender, race/ethnicity, and socioeconomic status. *Educational Researcher*, *47*(8), 525–531.

Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, *111*(28), 10107–10112.

Shahrill, M. (2013). Review of effective teacher questioning in mathematics classrooms. *International Journal of Humanities and Social Science*, *3*(17), 224–231.

Sigler, T., Martinus, K., & Loginova, J. (2021). Socio-spatial relations observed in the global city network of firms. *PLoS One*, *16*(8), e0255461.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*(2), 1–40.

Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, *36*(1), 99–153.

Springer, J. T., Hatcher, J., & Powell, A. (2018). High-impact practices: The call for a commitment to quality educational experiences and inclusive excellence. *Assessment Update*, *30*(4), 6–11.

Sundberg, M. D., & Moncada, G. J. (1994). Creating effective investigative laboratories for undergraduates. *BioScience*, *44*(1), 698–704.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483.

Todd, N. R., Blevins, E. J., & Yi, J. (2020). A social network analysis of friendship and spiritual support in a religious congregation. *American Journal of Community Psychology*, *65*(1-2), 107–124.

Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *Academy of Management Journal*, *28*(3), 712–723.

Tziner, A., Murphy, K., Cleveland, J. N., Yavo, A., & Hayoon, E. (2008). A new old question: Do contextual factors relate to rating behavior: An investigation with peer evaluations. *International Journal of Selection and Assessment*, *16*(1), 59–67.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS).. (2020). *IPEDS Fall 2010 through Fall 2019, Completions component*. Retrieved on December 6, 2022, from https://nces.ed.gov/programs/digest/d20/tables/dt20_318.45.asp

Van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social networks*, *31*(1), 52–62.

Vega Yon, G. G., Slaughter, A., & de la Haye, K. (2021). Exponential random graph models for little networks. *Social Networks*, *64*, 225–238.

Wei, C. A., & Woodin, T. (2011). Undergraduate research experiences in biology: Alternatives to the apprenticeship model. *CBE—Life Sciences Education*, *10*(2), 123–131.

Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social Networks*, *18*(4), 347–354.

Wells, J. E. (2019, July). Modeling student collaborations using valued ERGMs. In: *Proceedings of the Physics Education Research Conference (PERC)* (pp. 633–638).

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American journal of sociology*, *116*(2), 583–642.

Wright, E.R., & Pescosolido, B. A. (2002). Sorry, I forgot": The role of recall error in longitudinal personal network studies. In Levy, J. A., & Pescosolido, B.A. (Eds.), *Social Networks and Health* (Vol. 8, pp. 113–129). Bingley, UK: Emerald Group Publishing.

Xu, J. Y., Liu, T., Yang, L. T., Davison, M. L., & Liu, S. Y. (2019). Finding college student social networks by mining the records of student id transactions. *Symmetry*, *11*(3), 307.