

Investigating students' course performance by groups' gender compositions

Emily M. Smith and Kristine Callan

Department of Physics, Colorado School of Mines, Golden Colorado 80401

Instructors and researchers have been exploring the effects of students' groups in physics class on their experiences and course performance in a variety of ways. Here, we analyzed students' course performance based on the gender composition of the studio group they worked in throughout a full semester. Data were collected from one semester of two calculus-based introductory courses at an engineering university. Students regularly attended studio sections where they worked in groups of three in problem solving and hands-on activities. We explored students' exam and non-exam course performances based on the gender composition of their studio group and found little to no differences between students' course performances based on the gender composition of their groups. Including measures of test confidence and sense of belonging mitigated average differences in exam performance between men and women. We find no evidence to support particular gender compositions of studio groups for improving students' course performance.

I. INTRODUCTION

Effective group composition in active engagement physics classrooms has been a topic of discussion for at least the past three decades. Yet, studies find different outcomes, suggesting different classroom populations and environments may call for a variety of different techniques for approaching group compositions. However, some of these differences are due to the myriad of potential ways to define group composition, such as students' relative exam performance, incoming physics experience, gender, and preferences for group dynamics. This is in addition to differences in pedagogical approaches and many possible student outcomes to analyze (e.g., persistence in a course sequence, affective experiences, course performance). In this study, we explore students' exam and non-exam course performances based on the gender composition of their studio group. We find little to no differences between students' course performance based on the gender composition of their groups.

Some of the existing research on group composition in introductory physics classes focuses on groups' gender [1, 2] and physics ability [1–3] compositions and analyses a dimension of course performance (via exams or standardized diagnostics) as an outcome. For example, Heller and Hollabaugh proposed that cooperative groups are, on average, most effective with students with different incoming levels of physics ability (defined by individual exam scores) and that women are more successful in a course when grouped together [1]. Contrastingly, Callan *et al.* compared groups with heterogeneous and homogeneous physics abilities (defined by pre-instruction FMCE scores) and gender compositions and found no differences in students' post-instruction FMCE performance [2].

Other studies, especially in the introductory physics lab setting, find that students fall into gendered roles while working in groups. For example, Quinn *et al.* identified that men and women implicitly take up systematically different roles within their lab groups when a variety of roles are available within a group [4]. Similarly, Doucette *et al.* found that some roles that women tend to assume in mixed gendered groups may inhibit growth of students' physics identity [5]. These studies suggest that gender composition of groups may significantly change students' experiences in introductory physics classes.

In several biology courses, Ballen *et al.* demonstrated that exam and non-exam performance vary between men and women, with men, on average, outperforming women on exams and women outperforming men on non-exams [6]. The exam performance differences are mitigated by taking into account students' self-reported test anxiety. Studies have documented that women experience higher frequencies of test anxiety than men in a variety of undergraduate-level disciplines [7–9], which suggests that similar results to Ballen *et al.* may appear in introductory physics. However, Stang *et al.* found that in an introductory physics course test anxiety did not measurably contribute to students' final exam scores

when including a self-efficacy predictor [10], indicating that other attitudinal experiences than test anxiety may play a more prominent role in students' course performance.

Several studies highlight the importance of students' feeling of belonging in physics and university-level courses (e.g., Refs. [11–15]), and mens' and womens' sense of belonging in introductory physics courses may differ within the same course (e.g., Ref. [15]). For example Rainey *et al.* found that students who persist in a physics major report greater sense of belonging than those who choose to leave, with perceived competence as a contributor to belonging [13]. Students' affective experiences and performance in these courses often coincide with a strong sense of belonging in the course [11–15].

Therefore, in this study we also focus on gender composition of groups with the inclusion of measures of test confidence and sense of belonging because they have been shown to differently affect mens' and womens' course performance in science courses. We provide a preliminary example of the relationships between students' in-class group composition, self-reported testing confidence, sense of belonging, and course performance, separated by exam and non-exam performance. In doing so, we intend to highlight avenues for researchers to investigate decisions about group composition in introductory physics courses.

II. DATA AND CONTEXT

Data were collected from one semester of two calculus-based introductory courses in mechanics and electricity and magnetism (E&M) at Colorado School of Mines, an engineering university. In these courses, students attended (either remotely or in-person): (1) discussion sections where an instructor led the students through activities such as clicker questions, individual problem solving, and short discussions with peers and (2) studio sections where students worked in groups of three in group problem solving and hands-on activities. Studio groups were assigned at the beginning of the semester and maintained throughout the semester, except in the case of students withdrawing from the course. Students who opted to take the course fully remote were grouped with students who were attending in-person. Otherwise, groups were randomly assigned with no regard to student demographics or incoming physics experience.

As part of the course, students took surveys at the beginning and end of the course that included items intended to measure students' feelings of test confidence and belonging in the course. For example, an item pertaining to test confidence (when reverse scored) prompts students to rate their level of agreement with "I have an uneasy, upset feeling when I take a physics test." An item pertaining to sense of belonging is "I feel like I can be myself in this physics class." The items are part of a larger set of items developed by Singh's group and results from subsets of items have been presented in studies

TABLE I. All numbers are percentages; means with standard errors are weighted corresponding to the grading scheme. Extra credit is included in overall scores. Final exam was optional and reduced the weights of the three midterm exams. Students in studio groups with balanced gender compositions are excluded because there were no research consenting students in mechanics and only 13 students in E&M who were in such groups. All N in subsequent analyses are the same as those listed in this table.

Course	Gender	Group	Overall	Exam performance				Non-exam performance		
				Mean	Exam 1	Exam 2	Exam 3	Mean	Studio	Homework
Mechanics	Women	All (N=88)	84.4(1.0)	71.9(1.6)	72.3(1.8)	73.2(1.8)	70.3(1.9)	92.1(0.6)	91.7(0.5)	92.8(1.0)
		SG (N=32)	84.8(1.7)	71.8(2.8)	71.8(3.2)	71.5(3.5)	72.1(3.2)	92.5(0.8)	91.7(0.7)	93.7(1.3)
		MGW (N=50)	84.5(1.3)	72.5(2.1)	73.5(2.4)	74.6(2.3)	69.3(2.6)	92.4(0.9)	92.1(0.7)	92.8(1.5)
		MGM (N=6)	83.3(3.1)	68.0(3.4)	64.8(4.9)	70.8(4.5)	68.5(6.2)	87.6(3.2)	87.6(3.1)	87.6(4.5)
	Men	All (N=156)	86.5(0.7)	78.6(1.1)	80.5(1.4)	78.8(1.4)	76.5(1.4)	89.7(0.6)	89.9(0.5)	89.4(0.9)
		SG (N=124)	86.9(0.8)	79.1(1.3)	81.0(1.6)	79.0(1.6)	77.2(1.5)	90.1(0.6)	90.3(0.5)	89.8(1.0)
		MGW (N=25)	85.3(1.7)	76.4(2.6)	77.7(3.0)	77.0(3.3)	74.5(3.3)	89.1(1.5)	89.1(1.5)	89.2(2.6)
		MGM (N=7)	83.9(4.1)	77.5(5.8)	81.4(6.6)	80.9(5.6)	70.1(7.1)	84.1(4.6)	85.1(4.9)	82.5(5.8)
		E&M	All (N=162)	86.4(0.8)	76.4(1.1)	84.0(1.3)	75.3(1.2)	69.9(1.3)	94.2(0.7)	96.9(0.5)
Women	SG (N=34)	86.1(2.0)	76.2(2.6)	80.9(3.1)	76.4(2.7)	71.2(2.9)	94.1(1.5)	97.8(0.4)	88.5(2.9)	
	MGW (N=57)	86.2(1.2)	75.9(1.6)	84.7(2.0)	75.6(1.8)	67.5(2.2)	94.3(1.3)	96.6(1.0)	90.2(2.1)	
	MGM (N=64)	85.9(1.3)	75.8(1.7)	83.5(2.2)	73.7(2.0)	70.4(2.0)	93.7(1.1)	96.4(0.7)	89.4(1.9)	
	Men	All (N=225)	88.8(0.7)	82.6(0.9)	91.1(1.2)	81.0(1.0)	75.9(1.1)	91.1(0.6)	96.3(0.4)	83.9(1.1)
		SG (N=102)	88.7(0.9)	82.7(1.3)	89.9(1.8)	81.0(1.4)	77.1(1.6)	90.8(0.9)	96.1(0.4)	83.6(1.6)
MGW (N=28)		91.9(1.8)	84.6(2.7)	95.3(3.0)	79.6(3.4)	79.1(2.8)	95.5(0.8)	97.7(0.6)	91.5(1.6)	
MGM (N=89)		88.2(1.1)	81.8(1.4)	90.9(2.1)	81.6(1.3)	72.9(1.8)	90.6(1.0)	96.9(0.5)	82.0(2.0)	

such as Refs. [16–18]. We performed exploratory factor analyses on students’ responses to the items and found that items factored similarly to those found by Singh’s group (e.g., Ref. [16]). Therefore, we made no modifications to the items corresponding to test confidence and belonging. To score test confidence and belonging, we averaged students’ individual scores on items pertaining to those constructs. Each item had options ranging from strongly disagree (scored as 1) to strongly agree (scored as 5). We reverse scored test confidence items because agreement with the items were indicative of test anxiety rather than confidence. Possible scores for test confidence and belonging range from 1 to 5.

Demographic data were from the university registrar, to which students self-reported their demographic data. Unfortunately, the registrar only allows a binary selection of gender, so our analyses treat gender as a binary. In our analyses, we then used students’ gender relative to their studio group gender-composition as predictors: (1) single-gender (SG), (2) mixed gender with predominantly women (MGW), and (3) mixed gender with predominantly men (MGM). We excluded students from groups with gender-balanced composition due to small numbers of participants; there were no research consenting students in these groups in mechanics and only 6 men and 7 women in E&M.

We modeled exam performance and non-exam performance to distinguish whether differences occurred across

assessments of varying stakes. Exam performance is modeled using a repeated measures mixed model with random intercepts to account for variations in individuals’ exam performances. We excluded final exam performance from these analyses because both courses had an optional final exam, so most students opted out of the final exam. Non-exam performance is modeled using multiple linear regressions and is treated as an overall outcome. We express all results as scores out of 100% for ease of interpretation for the reader.

For both exam and non-exam performance outcomes, we tested four models. This study focuses on the effects of gendered group composition on students’ performance, so we consistently include gender relative to group composition in all models. We included combinations of test confidence and belonging in three of the four models because they are both closely tied to course performance outcomes in prior studies. These choices in models allows us to disentangle possible effects of group composition, test confidence, and sense of belonging for course performance in this population.

III. RESULTS

Table I shows descriptive statistics for students’ course performance with separations by gender and group composition. On average, women tended to outperform men on non-exam course components, however, men outperformed women on exams. Consistent trends based on

TABLE II. Test confidence and belonging by group composition. Possible scores range from 1 (low test confidence or sense of belonging) to 5 (high test confidence or sense of belonging).

Course	Gender	Group	Test Confidence	Belonging	
Mechanics	Women	All	2.5(0.1)	3.3(0.1)	
		SG	2.6(0.2)	3.3(0.2)	
		MGW	2.5(0.1)	3.3(0.2)	
		MGM	1.8(0.2)	2.9(0.4)	
		Men	All	3.5(0.1)	3.9(0.1)
			SG	3.6(0.1)	3.9(0.1)
	E&M	Women	All	2.6(0.1)	3.4(0.1)
			SG	2.4(0.2)	3.5(0.2)
			MGW	2.6(0.2)	3.4(0.1)
		Men	All	3.4(0.1)	3.8(0.1)
			SG	3.4(0.1)	3.7(0.1)
			MGM	3.4(0.1)	3.8(0.1)

group composition are not immediately apparent, which may indicate little correlation of group composition with course performance.

Table II shows the average scores of test confidence and belonging for the same populations of students. Across both courses, women systematically self-reported less test confidence, which implies higher frequencies of test anxiety compared to men. For sense of belonging, men tend to feel stronger belonging in the course than women. However, the difference in average belonging is less than that of test confidence. For our population, test confidence and belonging are substantially correlated (Pearson's coefficient of 0.68 for mechanics and 0.55 for E&M). Similar to course performance measures, there are no apparent trends that indicate group composition is correlated with test confidence or belonging.

A. Exam performance

Tables III and IV show predicted exam performance for mechanics and E&M, respectively. In both courses, the inclusion of the test confidence score mitigates gender differences, which agrees with results of analyses in biology education research [6]. However, a slight trend may persist in E&M for women to score lower than men even when test confidence is taken into account. Belonging appears to serve a similar role as test confidence, likely due to the relatively high correlation between the two. However, with the inclusion of belonging, in E&M the gender difference trend also remains but, again, appears to mostly mitigate differences in mechanics. When

TABLE III. Mechanics repeated measure mixed models predicting exam performance. Scores are reported out of 100%. Only 13 students were in MGM groups, so we also removed those students from this analysis. The base for comparison is men in single-gendered groups.

Predictors	Model 1	Model 2	Model 3	Model 4
(Intercept)	79.1(1.3) ^o	54.9(3.0) ^o	46.9(3.8) ^o	44.7(3.7) ^o
SG (Women)	-7.3(2.9)*	-0.6(2.6)	-2.3(2.6)	-0.2(2.5)
MGW (Women)	-6.6(2.5) [†]	1.2(2.3)	-1.4(2.2)	1.3(2.2)
MGW (Men)	-2.7(3.2)	0.3(2.8)	-2.2(2.8)	-0.6(2.7)
Test Confidence	-	6.7(0.8) ^o	-	3.9(1.0) ^o
Belonging	-	-	8.2(0.9) ^o	5.1(1.2) ^o
AIC	5735	5668	5665	5649
BIC	5762	5700	5697	5686

*p < 0.05; [†]p < 0.01; ^op < 0.001

TABLE IV. E&M repeated measure mixed models predicting exam performance. The base for comparison is men in single-gendered groups.

Predictors	Model 1	Model 2	Model 3	Model 4
(Intercept)	82.7(1.3) ^o	66.2(2.2) ^o	60.5(2.9) ^o	57.8(2.9) ^o
SG (Women)	-6.5(2.7)*	-1.7(2.5)	-5.4(2.5)*	-2.5(2.4)
MGW (Women)	-6.7(2.2) [†]	-3.2(2.1)	-4.8(2.1)*	-3.2(2.0)
MGM (Women)	-6.8(2.1) [†]	-3.2(2.0)	-4.2(2.0)*	-2.8(2.0)
MGW (Men)	2.0(2.9)	1.9(2.6)	2.3(2.6)	2.2(2.6)
MGM (Men)	-0.9(2.0)	-1.1(1.8)	-1.2(1.8)	-1.2(1.7)
Test Confidence	-	4.9(0.6) ^o	-	3.3(0.6) ^o
Belonging	-	-	6.0(0.7) ^o	3.7(0.8) ^o
AIC	9410	9341	9347	9322
BIC	9450	9386	9392	9372

*p < 0.05; [†]p < 0.01; ^op < 0.001

both test confidence and belonging measures are included (Model 4), results are similar to including only one of the measures. In these analyses, there is no indication that the gender composition of the studio groups correlates with students' exam performance.

B. Non-exam performance

Tables V and VI show students' performance on non-exam components of the courses, as weighted by course syllabi. There is a ceiling effect for scores across both courses (explaining the poor model fitting), which is expected with grades that are substantially participation-based. Women tended to perform better than men on non-exam components of the course with the exception of men in women-dominant groups in E&M. This effect remained even with the inclusion of students' sense of belonging in the course, suggesting that on average women outperform men on non-exam course components regardless of their group's composition. Perhaps expected, test confidence did not measurably affect students' non-exam performance in E&M. However, students' sense of belonging in the course consistently was a strong predic-

TABLE V. Mechanics linear regressions predicting non-exam performance. Scores are weighted and reported out of 100%.

Predictors	Model 1	Model 2	Model 3	Model 4
(Intercept)	90.1(0.6) ^o	84.1(1.4) ^o	83.9(1.8) ^o	83.2(1.8) ^o
SG (Women)	2.4(1.3)	4.0(1.3) [†]	3.3(1.3) [†]	4.1(1.3) [†]
MGW (Women)	2.3(1.1) [*]	4.2(1.1) ^o	3.3(1.1) [†]	4.2(1.1) ^o
MGW (Men)	-1.0(1.4)	-0.2(1.3)	-0.9(1.4)	-0.3(1.4)
Test Confidence	–	1.6(0.4) ^o	–	1.4(0.5) [†]
Belonging	–	–	1.6(0.5) ^o	0.5(0.6)
R_{adj}^2	0.02	0.10	0.07	0.10
AIC	1516	1499	1506	1500
BIC	1534	1519	1527	1524

*p < 0.05; †p < 0.01; °p < 0.001

TABLE VI. E&M linear regressions predicting non-exam performance. Scores are weighted and reported out of 100%.

Predictors	Model 1	Model 2	Model 3	Model 4
(Intercept)	90.8(0.9) ^o	89.0(1.6) ^o	83.2(2.0) ^o	83.6(2.1) ^o
SG (Women)	3.3(1.7)	3.8(1.8) [*]	3.7(1.7) [*]	3.3(1.8)
MGW (Women)	3.5(1.5) [*]	3.9(1.5) [†]	4.1(1.4) [†]	3.9(1.4) [†]
MGM (Women)	2.8(1.4) [*]	3.2(1.4) [*]	3.7(1.4) [†]	3.5(1.4) [*]
MGW (Men)	4.7(1.9) [*]	4.7(1.9) [*]	4.8(1.8) [†]	4.8(1.8) [†]
MGM (Men)	-0.2(1.3)	-0.3(1.3)	-0.3(1.2)	-0.3(1.2)
Test Confidence	–	0.5(0.4)	–	-0.5(0.5)
Belonging	–	–	2.0(0.5) ^o	2.4(0.6) ^o
R_{adj}^2	0.03	0.03	0.07	0.07
AIC	2692	2692	2677	2678
BIC	2720	2724	2709	2714

*p < 0.05; †p < 0.01; °p < 0.001

tor of non-exam performance in E&M. In mechanics, test confidence appears to play a measurable role, however, we suspect this may be due to the correlation between test confidence and belonging. We suspect that students who perform better on non-exam components have a network of support and/or collaborators within the course, which leads to a strong sense of belonging.

IV. DISCUSSION

In this study, we found no measurable differences in course performance according to the gender composition of students' studio groups. In both mechanics and E&M, men tended to outperform women on exams but women outperformed men on non-exam components of the course. Students' self-reported test confidence and sense of belonging substantially mitigated gender differences in exam performance, which suggests that the role of various experiences in physics classrooms affects students' performances in measurable ways. However, treating test confidence and sense of belonging as continuous variables may obscure some features of the data and may play a role in the null results.

Many of the course structures are intended to reduce test anxiety for students. In most semesters, students

take four exams with each weighted at about 15% of their grades. Collectively exams count for a substantial part of overall grades, however, each individual exam is relatively low stakes. Leading up to an exam, we provide many supports for studying including practice exams and multiple types of review sessions. However, we intend to explore additional ways to support students in reducing test anxiety in these courses, such as voluntary workshops that are tailored to our introductory physics courses.

Our courses are implicitly intended to promote students' sense of belonging by providing ample opportunities to interact with peers and instructors in both studio and discussion sections. However, we have not explicitly examined our course structures for building belonging. Neither course grades on a curve so that students are encouraged to collaborate when working through homework and studying for exams. Instructors regularly encourage students to collaborate, both inside and outside of the classroom environments, and to attend teaching assistant and faculty office hours. Both courses use a variety of assessment methods, many of which are low-stakes and based on in-class collaboration with peers. We hope to investigate how different course structures support or detract from students' sense of belonging to further develop course structures to support belonging.

Our work is limited by few participants who were in particular group compositions and by use of registrar data for identifying students' gender. For example, in mechanics there were very few participants in MGM groups. Due to the typical three-person group size, we also have very few to no groups that were mixed gender in equal proportions. Additionally, the survey items for test confidence and sense of belonging were administered only at the end of the semester after the exams. Students' experiences could fluctuate throughout the semester depending on many influences such as course content, how studio group members are engaging, and support within the course as a whole. Lastly, the study takes place during the COVID-19 pandemic, so the course structure is quite different and relied on studio groups to communicate to each other across a video platform. Developing effective group dynamics in this environment may require new sets of skills that college students and instructors had not yet had to develop compared to most prior studies within physics education research.

Developing effective grouping in courses with collaborative components remains an open avenue for investigations in physics education research. Here, we find little evidence to support particular gender compositions of studio groups for improving students' course performance. However, much of students' experience in a course may be due to their interactions with their peers, especially those that they frequently work with in-class, and may not directly affect performance outcomes. Simultaneous investigations of students' course performance and affective experiences may begin to provide insight to how various group compositions affect students' experiences in physics.

-
- [1] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. part 2: Designing problems and structuring groups, *American journal of physics* **60**, 637 (1992).
- [2] K. E. Callan, B. Wilcox, and W. Adams, Testing group composition within a studio learning environment, in *Physics Education Research Conference 2017*, PER Conference (Cincinnati, OH, 2017) pp. 72–75.
- [3] J. J. B. Harlow, D. M. Harrison, and A. Meyertholen, Effective student teams for collaborative learning in an introductory university physics course, *Phys. Rev. Phys. Educ. Res.* **12**, 010138 (2016).
- [4] K. N. Quinn, M. M. Kelley, K. L. McGill, E. M. Smith, Z. Whipps, and N. G. Holmes, Group roles in unstructured labs show inequitable gender divide, *Phys. Rev. Phys. Educ. Res.* **16**, 010129 (2020).
- [5] D. Doucette, R. Clark, and C. Singh, Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning, *Eur. J. Phys.* **41**, 035702 (2020).
- [6] C. J. Ballen, S. Salehi, and S. Cotner, Exams disadvantage women in introductory biology, *PLOS ONE* **12**, 1 (2017).
- [7] N. von der Embse, D. Jester, D. Roy, and J. Post, Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review, *Journal of Affective Disorders* **227**, 483 (2018).
- [8] M. S. Chapell, Z. B. Blanding, M. E. Silverstein, M. Takahashi, B. Newman, A. Gubi, , and N. McCann, Test anxiety and academic performance in undergraduate and graduate students, *Journal of Educational Psychology* **97**, 268 (2005).
- [9] M. I. Núñez-Peña, M. Suárez-Pellicioni, and R. Bono, Gender differences in test anxiety and their impact on higher education students' academic achievement, *Procedia - Social and Behavioral Sciences* **228**, 154 (2016), 2nd International Conference on Higher Education Advances, HEAd'16, 21-23 June 2016, València, Spain.
- [10] J. B. Stang, E. Altieri, J. Ives, and P. J. Dubois, Exploring the contributions of self-efficacy and test anxiety to gender differences in assessments, in *Physics Education Research Conference 2020*, PER Conference (Virtual Conference, 2020) pp. 497–502.
- [11] T. M. Freeman, L. H. Anderman, and J. M. Jensen, Sense of belonging in college freshmen at the classroom and campus levels, *The Journal of Experimental Education* **75**, 203 (2007), <https://doi.org/10.3200/JEXE.75.3.203-220>.
- [12] K. L. Lewis, J. G. Stout, S. J. Pollock, N. D. Finkelstein, and T. A. Ito, Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020110 (2016).
- [13] K. Rainey, M. Dancy, R. Mickelson, E. Stearns, and S. Moller, Race and gender differences in how sense of belonging influences decisions to major in stem, *International Journal of STEM Education* **5**, 1 (2018).
- [14] C. Good, A. Rattan, and C. S. Dweck, Why do women opt out? sense of belonging and women's representation in mathematics, *Journal of personality and social psychology* **102**, 700 (2012).
- [15] J. Stout, T. A. Ito, N. Finkelstein, and S. Pollock, How a gender gap in belonging contributes to the gender gap in physics participation, in *Physics Education Research Conference 2012*, PER Conference, Vol. 1513 (Philadelphia, PA, 2012) pp. 402–405.
- [16] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way, *Phys. Rev. Phys. Educ. Res.* **15**, 020148 (2019).
- [17] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, *Phys. Rev. Phys. Educ. Res.* **15**, 020119 (2019).
- [18] E. M. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with a's have similar physics self-efficacy as male students with c's in introductory courses: A cause for alarm?, *Phys. Rev. Phys. Educ. Res.* **14**, 020123 (2018).