

## Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?

Emily M. Marshman,<sup>1</sup> Z. Yasemin Kalender,<sup>2</sup> Timothy Nokes-Malach,<sup>3</sup>  
Christian Schunn,<sup>3</sup> and Chandralekha Singh<sup>2</sup>

<sup>1</sup>*Department of Physics, Community College of Allegheny County, Pittsburgh, Pennsylvania 15212, USA*

<sup>2</sup>*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*

<sup>3</sup>*Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*



(Received 26 June 2018; published 6 December 2018)

Self-efficacy can affect performance, career goals, and persistence. Prior studies show that female students have lower self-efficacy than male students in various science, technology, engineering, and mathematics (STEM) domains, and the self-efficacy gap is a factor that contributes to the low representation of female students in STEM. However, prior research has not decoupled self-efficacy differences from performance differences. This study examines the self-efficacy of male and female students with similar performance in introductory physics courses and investigates whether gender gaps in self-efficacy are persistent across different instructors and course formats. Students filled out a self-efficacy in physics survey before physics 1, before physics 2, and at the end of physics 2. Students' achievement was measured by their performance on research-based conceptual physics tests and course grades. The physics courses were taught by several instructors and varied in the type of pedagogy used, with some using a "flipped" format and others using a traditional, lecture-based format. We found that female students had lower self-efficacy than male students at all performance levels in both physics 1 and physics 2. The self-efficacy gaps continued to grow throughout the introductory physics course sequence, regardless of course format (i.e., traditional or flipped) and instructor. The findings suggest that female students' self-efficacy was negatively impacted by their experiences in introductory physics courses, and this result is persistent across various instructors and course formats. Female students' lower self-efficacy compared to similarly performing male students can result in detrimental short-term and long-term impacts.

DOI: [10.1103/PhysRevPhysEducRes.14.020123](https://doi.org/10.1103/PhysRevPhysEducRes.14.020123)

### I. INTRODUCTION

#### A. Underrepresentation of women in STEM fields

In many countries, there has been much focus on the underrepresentation of women in engineering and physical science fields, from elementary to high school, at the undergraduate level, in career fields, and in leadership positions. In the United States, the numbers of women studying science, technology, engineering, and mathematics (STEM) and pursuing STEM careers have not changed significantly in the last decade [1]. For example, since 2000, women have earned approximately 20% of the bachelor's degrees awarded in physics and engineering, with a similar underrepresentation at the masters and Ph.D. levels [2–4]. Some efforts have been made to increase the diversity in STEM courses and STEM occupations, yet the

reasons for the low percentage of women in science and engineering are not fully understood and progress has been slow (e.g., less than 1% change over the period from 2006 to 2014).

#### B. Self-efficacy and engagement in STEM fields

Deciding to pursue a STEM career and continuing a pathway to physical sciences and engineering is generally affected by several interrelated factors, including students' prior preparation and skills [5–16], quality of teaching and type of teaching approach [17–21], sociocultural factors [22–35], and motivational factors [36–49]. Motivational factors, such as intelligence mindset, interest in science, and self-efficacy in science, can also influence students' decisions to major and persist in STEM fields [36–38]. In particular, self-efficacy is the belief in one's capability to be successful in a particular task, course, or subject area [40–43] and it is one aspect of motivation. Self-efficacy can impact one's interests [44]. Because self-efficacy can shape interest, it also influences engagement during learning [45]. Furthermore, students with high self-efficacy in a domain often enroll in more difficult courses in that domain than

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

those with low self-efficacy because they perceive difficult tasks as challenges rather than threats [46]. Self-efficacy in STEM also predicts initial career goals and enrollment in STEM courses [37,43,47] as well as persistence toward long-term career goals [44,48].

Self-efficacy in particular can have large long-term effects because it can thrust a student into a feedback loop which can impact students' self-efficacy and performance in a positive or negative way. As students complete short-term goals (e.g., complete a physics course requirement for a physical science or engineering major), they obtain feedback about their performance that shapes their self-efficacy. More surprisingly, research suggests that the reverse effect also occurs: for a variety of reasons, self-efficacy beliefs constrain performance in science courses beyond the more normative effects of students' prior knowledge and skills [49]. For example, students with a high self-efficacy are more likely to exhibit effective learning strategies such as self-monitoring [44,50,51] and tend to make more efficient use of problem-solving strategies and time management [44]. In addition, students with high self-efficacy are less likely to reject correct hypotheses prematurely and tend to be better at solving conceptual problems than students with low self-efficacy but equal performance [52]. As a result, there are feedback loops in which higher initial self-efficacy produces higher performance which further strengthens self-efficacy; by contrast, lower initial self-efficacy produces lower performance which further weakens self-efficacy.

Unfortunately, several studies have shown that female students have significantly lower self-efficacy than male students in STEM-related domains [23,53,54] including in mathematics, engineering [14,24,55], and computer science [56,57]. In physics, female students report significantly lower self-efficacy than male students [58–62], even in interactive engagement courses [60] and even if female students had received prior physics instruction in high school [61]. Furthermore, female students experience a decline in self-efficacy throughout their engineering education at college [55]. A strong sense of self-efficacy, especially for female students in physical science and engineering courses, can help them persist in STEM fields [63–67]. Given the central role of self-efficacy in career choice and career persistence, the self-efficacy gap likely contributes to the low representation of women pursuing careers in physical science and engineering fields [21]. Therefore, understanding the source and nature of the gap is important.

### C. Rationale for the study and research questions

Prior studies have mainly focused on using self-efficacy to predict performance outcomes [44,49,58]. However, this study aims to investigate whether there are differences in the self-efficacy of male and female students throughout introductory physics courses at matched performance

levels. In particular, it is important to investigate whether self-efficacy is a reflection of students' actual performance, or whether there is a self-efficacy gender gap over and above performance that may be due to environmental factors such as stereotype threat. There are two types of performance that we look at: conceptual learning test results and students' grades. Differences in self-efficacy even within similarly performing male and female students can have many detrimental effects. A large underestimate of one's capability and/or performance can impact interest and goals. For example, if female students inaccurately perceive that they are not capable of succeeding in a STEM field, that could lead to decreased interest in STEM disciplines and underrepresentation of women in STEM fields.

In mathematics, prior research has shown that male students assess their own mathematical ability more favorably than female students of similar ability in high school [68]. Male students were more likely than equally performing female students to enroll in calculus courses in high school, and this difference was shown to be due to differences in male and female students' self-efficacy [68]. Since enrolling in calculus in high school has a large influence on the decision of women to choose a STEM major, female students' self-efficacy in mathematics is correlated with whether they enroll in advanced mathematics courses in high school and whether they decide to major in a STEM discipline [68]. Research also suggests that many women believe that they must achieve at exceptionally high levels in mathematics and science to be successful STEM professionals [69]. Interviews with and surveys of women in engineering programs suggest that the exit of women from engineering programs is not driven primarily by their performance or success, but partly because women have low self-efficacy and negatively interpret their grades [38]. For example, female students may view a "B" grade as a poor performance even though a "B" grade is above average [68]. Furthermore, research suggests that women who are equally successful as men at a mathematical task are less likely to compete at the task at the same rate as male students [70]. Much less research has examined whether self-efficacy in science has a similarly gendered misperception of ability.

The most common dropout points for women from physical science or engineering pathways are during the first and second years of college [38]. Since college level physics is a key prerequisite to obtaining a physical science or engineering degree and students usually take physics sequences in their first year, it is important to examine the self-efficacy of similarly performing male and female students in college level physics courses for scientists and engineers. However, research has not systematically examined the self-efficacy of male and female students who have similar performance outcomes in physics. It may be that female students simply have lower performance levels

and lower self-efficacy that reflects the lower performance levels; alternatively, it may be that female students' self-efficacy in physics is below similarly performing male students in college level physics courses. Women who have high standards for achievement in physics may drop out of physics courses and leave a STEM major at a higher rate than male students if they underestimate their own capability to succeed in physics. Thus, we systematically investigated the self-efficacy of similarly performing male and female students in the context of the typical two-semester physics course sequence for engineering and physical science majors.

Matching students by performance requires selecting a common performance metric. Student performance can be assessed in many different ways, and there can be gender biases in particular assessment formats [71–77]. For example, grades based upon participation or explanation quality in open-ended responses may reflect biases in the evaluator or rater. Therefore, using standardized research-based assessments provides one strong method for examining this topic. However, performance shapes self-efficacy through explicit feedback, and so the grades that students obtain in courses provide another important performance metric for this topic. Therefore, we examine self-efficacy by gender in two different ways: once matched by research-based standardized conceptual physics assessments and again matched by course grades.

Since prior research has suggested that instructor attitudes, assessment practices, and course format can differentially influence learning, attitudes, and retention in STEM courses [19,28,29,30,78,79], we also examine students' self-efficacy across different course types (e.g., flipped vs traditional format) and different instructors. In particular, it has been found in some research studies that students' self-efficacy is impacted by different teaching methodologies [58,60,61,80–84]. Our investigation can shed light on whether self-efficacy gender differences at matched performance levels are broadly generalizable across different instructional contexts.

In particular, our research questions (RQ) are as follows:

**RQ1.** What is the self-efficacy of female and male students throughout a two-semester physics course sequence when prior knowledge differences on physics conceptual surveys are accounted for?

**RQ2.** What is the self-efficacy of female and male students throughout a two-semester physics course sequence when students' course performance is accounted for?

**RQ3.** Are the effects consistent across different instructors and different types of physics courses, e.g., flipped vs traditional formats in physics 1 and physics 2?

## II. METHODOLOGY

To investigate the self-efficacy of introductory students with similar performance outcomes in physics,

we administered a motivation survey and physics conceptual assessments in several sections of a two-course calculus-based introductory physics sequence. We collected data across two consecutive academic years.

### A. Participants and class context

Participants were students enrolled in 9 sections of physics 1 and 11 sections of physics 2. These two large introductory physics courses are arranged in a two-semester sequence at the University of Pittsburgh, a large R1 public university. This calculus-based physics sequence is typically taken by engineering and physical science majors as a requirement. Physics 1 includes topics such as kinematics, forces, energy and work, rotational motion, gravitation, and oscillations and waves. Physics 2 includes topics such as electricity and magnetism, electromagnetic waves, reflection, interference, and diffraction. Both courses included four lecture hours and one recitation hour per week. The recitations were mandatory and included a weekly, low stakes quiz. Most of the sections of both courses were taught in a traditional, lecture-based format, but there were four sections of physics 1 and four sections of physics 2 that were taught in a “flipped” format [85,86]: students watched lecture videos before attending the lectures, during which they worked on collaborative group problem solving and clicker questions (in which the students answered questions individually, discussed their answers with a peer, and then answered the questions again).

Table I shows the number of students in introductory physics courses who completed the motivation survey and/or physics conceptual survey throughout the introductory physics course sequence. The number of students is different at different points of time because some students did not take physics 2 and some students were not present in the lecture or recitation section in which the surveys were administered (however, when the analysis is performed again for matched students, the results are qualitatively similar).

The cohort of both courses was approximately 32% female. Seventy-five percent of the students were white, 13% of the students were Asian, and 12% of the students were black, Hispanic, or multiracial. Most of the students were between 18 and 19 years old. Sixty-eight percent of

TABLE I. The number of students completing the self-efficacy survey and the physics conceptual assessments at different points in time.

	Motivation survey		Conceptual Assessment	
	Pre	Post	Pre	Post
Physics 1	$N = 1054$	Data not collected	$N = 726$	$N = 644$
Physics 2	$N = 914$	$N = 630$	$N = 845$	$N = 807$

TABLE II. The physics self-efficacy survey. One item indicated with an (*R*) is reverse coded.

Survey item	Response options
1. I can complete the physics activities I get in a lab class	<input type="radio"/> Rarely <input type="radio"/> Half of the time <input type="radio"/> Most of the time <input type="radio"/> All of the time
2. If I went to a museum, I could figure out what is being shown about physics in [see options to the right]:	a. None of it b. A few areas c. Most areas d. All areas
3. I am often able to help my classmates with physics in the laboratory or in recitation.	
4. I get a sinking feeling when I think of trying to tackle difficult physics problems. ( <i>R</i> )	a. No! b. no
5. If I wanted to, I could be good at doing physics research.	c. yes d. Yes!
6. If I study, I will do well on a physics test.	

the students were enrolled in an engineering track, and the rest were science majors or other majors requiring physics.

### B. Validity and reliability of survey

The motivation survey was adapted and validated based upon previously developed survey instruments [82,87–89] and used in prior research studies [62,90,91]. The survey includes questions focused on several aspects of motivation, including interest and value associated with physics, intelligence mindset, and self-efficacy. In this paper, we focus only on students' responses to the self-efficacy questions. Table II shows the physics self-efficacy survey items, which all involved 4-point Likert scales. We note that the scale for some of the self-efficacy questions was in the form of "NO!, no, yes, YES!" This response scale has been extensively validated in prior studies, including studies of self-efficacy [54]. Our findings are similar to the validation of these rating scales in prior investigation [54]. In particular, this scale was used as opposed to "strongly disagree, disagree, agree, strongly agree" because students interpret these rating scales appropriately and because it reduces students' cognitive load, which is especially important for non-native speakers and for questions that ask raters to consider subtle differences in survey items [54]. The four-point scale also allows printing the scale next to each item so students do not need to search the instructions or map back to the initial scale as some surveys do, thereby further reducing potential errors in mapping or misremembering scale numbers. The physics self-efficacy survey items represent a diverse set of contexts and forms of interaction with physics and are meant to capture a general sense of efficacy with physics content. The item types are purposely varied to encourage processing of each item, rather than quickly responding similarly to each item.

Internal coherence and discriminability from other motivational constructs (e.g., identity, interest, and valuing of physics) was established through factor analysis, and item separation is shown in detail in our previous work [62]. We measured Cronbach's alpha for each construct to check internal consistency of the questions. The Cronbach's alpha is above 0.70 for all of the constructs and for self-efficacy questions it is 0.74, which is considered adequate [91,92]. Individual interviews with students also provided useful feedback for refining or discarding survey items.

We also ensured content validity, i.e., the degree to which the survey items reflect the domain of interest (in our case, self-efficacy), by taking steps to ensure that the respondents interpreted the survey questions as was intended. We conducted one-on-one interviews with 12 students (8 male and 4 female students) in introductory, calculus-based physics courses and 3 students in graduate-level physics courses (2 male and 1 female students) using a think-aloud protocol [93] to verify that the students interpreted each question as intended. Students voluntarily participated in the interviews and were compensated for their time. Most of the students enrolled in the introductory physics courses planned to major in engineering or physical science. The interviewed students' responses were audio recorded. Each interview took approximately an hour. During the individual interviews, students were asked to read each question aloud and explain how they interpreted the question. They were also asked to respond to the survey questions and give an explanation for their responses. During the interviews and discussions, we paid attention to respondents' interpretations of the questions and modified them accordingly in order to clarify their intent. We note that none of the interviewees had difficulty interpreting the "NO!, no, yes, YES!" response scale.

TABLE III. The cutoff scores (in assessment percentages) for each FCI pretest or post-test and CSEM pretest or post-test performance bins, along with percentage of female students within each bin.

Assessment	Time point	Bin cutoff scores	% Female
FCI (physics 1; 33% female)	Pretest $N = 726$	Low (0%–45%)	52%
		Medium (46%–67%)	33%
		High (68%–100%)	14%
	Post-test $N = 644$	Low (0%–56%)	51%
		Medium (57%–81%)	33%
		High (82%–100%)	15%
CSEM (physics 2; 32% female)	Pretest $N = 845$	Low (0%–33%)	48%
		Medium (34%–44%)	26%
		High (45%–100%)	22%
	Post-test $N = 807$	Low (0%–42%)	44%
		Medium (43%–63%)	30%
		High (64%–100%)	26%

### C. Measures

**Conceptual physics knowledge:** In physics 1, the Force Concept Inventory [94] was administered. In physics 2, the Conceptual Survey of Electricity and Magnetism [95] was administered. Both assessments have been extensively validated as measures of conceptual understanding of core physics phenomena and principles within each of the two topic areas [94,95]. They also correspond to the earlier parts of the course that would be covered by all sections; coverage of later course content often varies across sections (e.g., optics topics are not consistently covered in physics 2).

For a number of the analyses, we grouped the students by performance into three equal-width performance bins (i.e., the number of students in each bin was approximately the same). Since students learned content from pre to post, the cutoff scores were determined relative to the time point. See Table III for the cutoff scores for each bin at each time point.

**Course grades:** We also obtained students' final grades in each course. Students' final grades are mostly based on high-stakes assessments that may have more impact on their self-efficacy than their performance on the physics conceptual surveys. Since some grades occur infrequently, for analysis, students were grouped into bins by grade ranges: (i) C– and below (considered insufficient to move on to the next course); (ii) C and C+; (iii) B–, B, B+; and

(iv) A–, A. Table IV shows the grade distribution for students' final grades in each course, along with the percentage of students in each bin who are female.

### D. Procedures

The self-efficacy survey was given at three time points: at the beginning of physics 1, the beginning of physics 2, and the end of physics 2; the end of physics 1 was excluded to avoid survey fatigue or redundancy with the beginning of physics 2 only a few weeks later. The relevant physics conceptual assessment was given as a pretest or post-test in each course. The self-efficacy survey and physics conceptual assessments were typically administered in the first and last recitations of the course, although one instructor chose to give the survey in the lecture portion of the course. In the first year of administration, some instructors chose to give the survey and assessments in a written format, whereas others chose to use an online format completed outside of class. We found that the participation rates were significantly lower for students who were given the online format. Thus, in subsequent administrations, only the written format was used. The self-efficacy survey was completed by most students in a couple of minutes (embedded in a larger motivational survey taking between 10 and 15 min), and the students worked through the conceptual physics assessments in the remaining class time (approximately 35–40 min).

TABLE IV. The letter grade distributions for students' final grades in each course, including % female within each bin.

Course	Total	C– or below	C, C+	B–, B, B+	A–, A
Physics 1	$N = 1130$	$N = 210$	$N = 375$	$N = 368$	$N = 177$
	33% female	34% female	44% female	33% female	24% female
Physics 2	$N = 1059$	$N = 195$	$N = 308$	$N = 384$	$N = 172$
	32% female	30% female	35% female	34% female	22% female

Instructors were encouraged to give a small amount of course credit to students for completing the surveys. The instructor or teaching assistant responsible for giving the survey was given the following script to announce before administering the survey to the students to encourage students to take the assessments seriously: “We are surveying you on your understanding and beliefs about physics in order to improve the class. Your responses will not be evaluated for grades except to make sure the responses were done seriously, rather than randomly.” The university provided course grades and student demographics (gender, ethnicity, and major), and this information was linked to the conceptual assessments and self-efficacy surveys through an honest broker; that is, the researchers only had access to the linked data in a de-identified form. We removed a small number of students whose gender was unidentified.

### E. Analysis

To determine whether there are differences in the self-efficacy of female and male students (overall, combining different instructors and across flipped vs traditional formats) controlling for performance on physics conceptual surveys, we performed a linear regression in which the dependent variable was students’ average post self-efficacy score in either physics 1 or physics 2 and the independent variables were gender and the students’ FCI or CSEM posttest scores. This analysis was also conducted separately for each instructor and format of physics course (i.e., flipped or traditional lecture-style format). This by-section analysis allowed us to determine whether gender had a significant impact on students’ end of semester self-efficacy (controlling for their post FCI or CSEM scores) across various instructors and course formats.

To take into account possible nonlinear effects, we also analyzed students’ mean self-efficacy scores by gender and conceptual survey performance bins. For each bin, we calculated the effect sizes [Cohen’s  $d = (\mu_1 - \mu_2)/\sigma_{\text{pooled}}$ , where  $\mu_1$  and  $\mu_2$  are male and female students’ average self-efficacy scores and  $\sigma$  refers to the pooled standard deviation] [96] between male and female students’ average self-efficacy scores to compare the two groups’ standardized means and determined whether the differences were significant using a two-way ANOVA. The ANOVA analysis was repeated across pretest and post-test data, in each case analyzing the relationship of conceptual assessment to its most temporally proximal self-efficacy data.

The same analysis was conducted using grade bins rather than conceptual assessment bins. In this analysis, we focused only on self-efficacy differences between male and female students who passed the course (i.e., received a grade of C or better in the course since a C or better is required to remain on the engineering or physical science track). For each final letter grade, we calculated the effect size between male and female students’ average self-efficacy scores and we checked whether the differences

were significant using a two-way ANOVA. We also repeated this analysis at the level of individual self-efficacy survey items to ensure that the gender differences in self-efficacy were robust to physics self-efficacy more broadly, rather than just efficacy about a particular aspect of physics (e.g., test taking).

## III. RESULTS

### A. Gender differences in self-efficacy, controlling for performance on standardized conceptual physics tests

In regards to research question 1, at the beginning of physics 1, a two-way ANOVA (self-efficacy is the outcome, gender and conceptual test bin are the factors) revealed a small but significant gender difference [ $F(1690) = 8.25$ ,  $p < 0.01$ ], a significant effect of FCI bin (higher performing students had higher confidence levels,  $p < 0.01$ ), and no significant interaction effect between gender and FCI bin. However, as shown in Table V and Fig. 1 (upper left), male students’ self-efficacy was higher than female students’ primarily in the low and medium FCI performance groups. The effect size differences in female and male students’ self-efficacy for the medium and low pre-FCI bins were greater than 0.30 (but less than 0.50); which is considered a medium effect size. To contextualize the effect size, female students in the medium FCI group had the same self-efficacy as male students in the low FCI group.

At the end of physics 1, the main effect of gender was also statistically significant [ $F(1318) = 22.28$ ,  $p < 0.001$ ] and even larger overall, and the interaction of gender and FCI performance bin was statistically significant [ $F(2317) = 3.76$ ,  $p < 0.05$ ]. As shown in Table V and Fig. 1 (upper right), the gender effect size was largest in the low group, moderate in the medium group, and approaching large in the high group. Interestingly for male students, self-efficacy was the same in the medium and low group (i.e., their self-efficacy was not influenced by relative performance levels). As a result of both patterns, the top performing female students had roughly the same self-efficacy as the lowest performing male students.

At the beginning of physics 2, both main effects of gender and CSEM performance bin were statistically significant [for gender,  $F(1777) = 49.44$ ,  $p < 0.001$ ; for CSEM bin,  $F(2776) = 60.70$ ,  $p < 0.001$ ], with no significant interaction effect between gender and CSEM bin. As shown in Table VI and Fig. 1 (lower left), the gender effect size was generally large (greater than 0.50), with female students in the medium CSEM group showing similar self-efficacy scores as male students in the low CSEM group, and female students in the high CSEM group having equal self-efficacy with male students in the medium group.

By the end of physics 2, both main effects of gender and CSEM performance bin were statistically significant [for

TABLE V. Self-efficacy scores in physics 1 binned by FCI scores (high, medium, low) for female and male students ( $M$  = mean,  $SD$  = standard deviation,  $N$  = number of students), along with Cohen's  $d$  for the gender contrast in means.

	Self-efficacy						FCI physics 1											
	Low						Medium						High					
	Pre			Post			Pre			Post			Pre			Post		
	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$
Female	2.58	0.47	122	2.30	0.54	46	2.76	0.35	68	2.65	0.32	32	3.04	0.38	34	2.83	0.43	20
Male	2.73	0.39	112	2.78	0.39	38	2.91	0.41	139	2.77	0.43	68	3.05	0.37	216	3.04	0.44	115
$d$ size	0.33 <sup>b</sup>	...	...	0.90 <sup>c</sup>	...	...	0.37 <sup>a</sup>	...	...	0.30	...	...	0.03	...	...	0.48 <sup>a</sup>	...	...

$d$  sizes are Cohen's  $d$  values between female and male students.

<sup>a</sup>Significant at the 0.05 probability level.

<sup>b</sup>Significant at the 0.01 probability level.

<sup>c</sup>Significant at the 0.001 probability level.

gender,  $F(1594) = 71.94$ ,  $p < 0.001$ ; for CSEM bin,  $F(2593) = 27.00$ ,  $p < 0.001$ ], with no significant interaction effect between gender and CSEM bin. In other words, students who earned higher scores in CSEM

reported higher self-efficacy. As shown in Table VI and Fig. 1 (lower right), the gender differences become even larger for the high CSEM performance group, increasing from  $d = 0.31$  to  $d = 0.78$  (which is considered a large

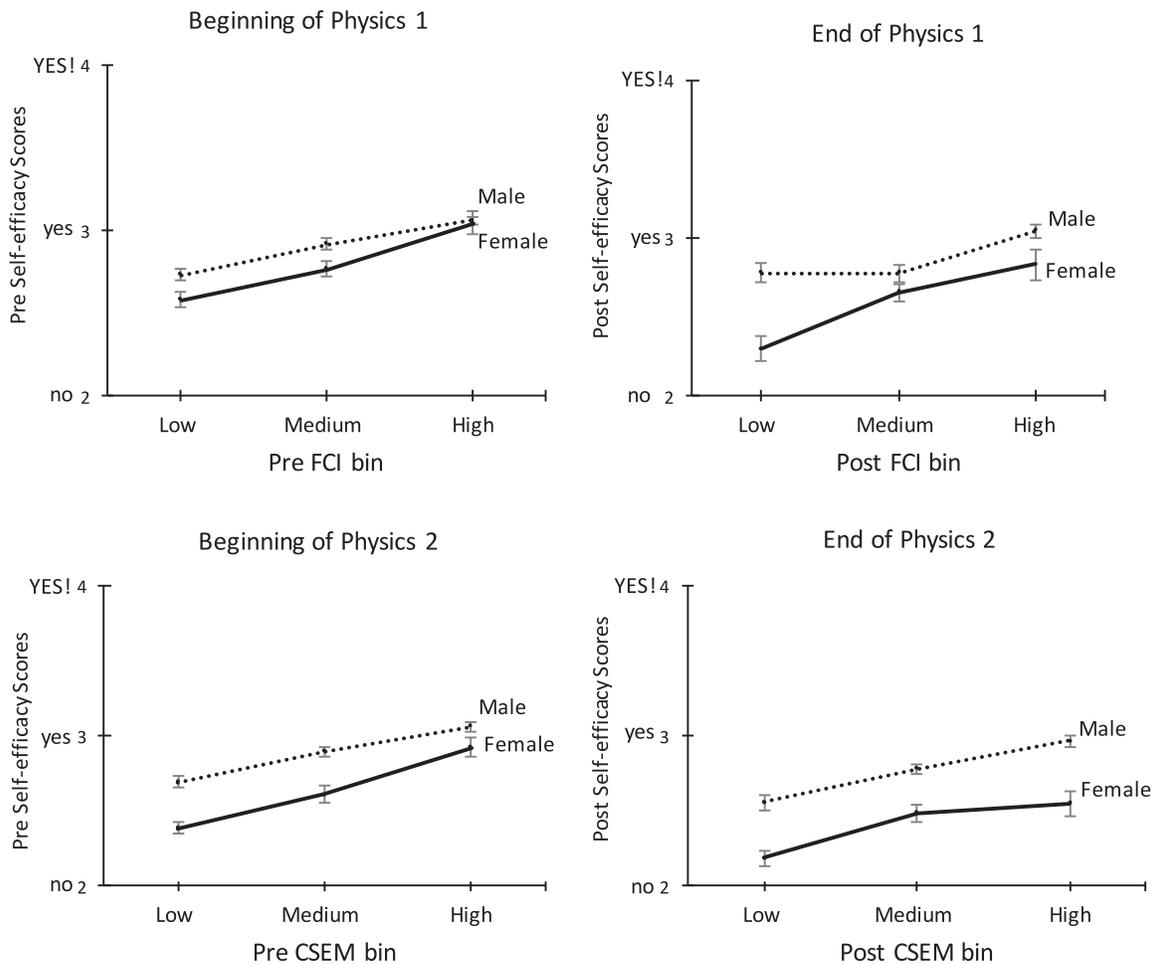


FIG. 1. Pre-self-efficacy (left column) and post-self-efficacy (right column) scores of female and male students binned by temporally proximal conceptual test score in physics 1 (top row) and physics 2 (bottom row) courses. Error bars represent standard error of the mean.

TABLE VI. Self-efficacy scores in physics 2 binned by CSEM scores (high, medium, low) for female and male students ( $M$  = mean,  $SD$  = standard deviation,  $N$  = number of students), along with Cohen’s  $d$  for the gender contrast in means.

	Self-efficacy									CSEM physics 2								
	Low			Medium			High			Medium			High					
	Pre			Post			Pre			Post			Pre			Post		
	$M$	$SD$	$N$															
Female	2.38	0.45	135	2.18	0.45	79	2.60	0.45	83	2.48	0.49	75	2.92	0.46	61	2.54	0.56	47
Male	2.69	0.45	137	2.56	0.47	90	2.90	0.40	182	2.77	0.45	154	3.06	0.43	200	2.96	0.47	150
$d$ size	0.65 <sup>b</sup>	...	...	0.74 <sup>b</sup>	...	...	0.66 <sup>b</sup>	...	...	0.61 <sup>b</sup>	...	...	0.31 <sup>a</sup>	...	...	0.78 <sup>b</sup>	...	...

$d$  sizes are Cohen’s  $d$  values between female and male students.

<sup>a</sup>Significant at the 0.05 probability level.

<sup>b</sup>Significant at the 0.001 probability level.

effect size). Importantly, at the end of this physics sequence, female students in the high CSEM group report similar post self-efficacy as male students in the low CSEM group.

**B. Gender differences in self-efficacy controlling for physics course grade**

In regards to research question 2 (self-efficacy by grade level), a two-way ANOVA (where self-efficacy is the outcome, gender and grade bin C, B, A are the factors) revealed that the differences in male and female students’ self-efficacy were statistically significant [ $F(1608) = 66.31, p < 0.001$ ], and the effect size was large (greater than 0.50; see Table VII and Fig. 2, top) in physics 1. There was also a main effect of course grade [ $F(2607) = 43.06, p < 0.001$ ] and no interaction effect between gender and grade [ $F(2607) = 1.65, p = 0.19$ ]. Female students had significantly lower self-efficacy compared to their male counterparts in all grade groups (A, B, C). Moreover, as the top of Fig. 2 shows, female students receiving A’s have similar self-efficacy as male students receiving C’s in physics 1.

Similarly, the gender gap in self-efficacy controlling for final course grade continues to persist in physics 2 with equally large effect sizes. The ANOVA had significant effects of gender [ $F(1555) = 52.44, p < 0.001$ ] and course grade [ $F(2554) = 21.32, p < 0.001$ ], and no interaction [ $F(2554) = 1.14, p = 0.32$ ]. Students who earned higher grades reported higher self-efficacy. In other words, the effect of gender on self-efficacy does not change based on the course grade. Female students had much lower self-efficacy than did male counterparts at each matched letter grade group and the gender effect sizes are all large (greater than 0.50). Again, women receiving A’s had similar self-efficacy scores to men receiving C’s in physics 2 (see Table VIII and Fig. 2).

In order to explore what fraction of the self-efficacy gender gap was related to students’ course performance versus students’ biased perception, we first measured students’ raw post self-efficacy differences between female and male students for physics 1 and physics 2, which are 0.39 ( $p < 0.001$ ) and 0.40 ( $p < 0.001$ ), respectively, and which favors the male students. Next, we controlled for students’ grade in physics 1 and physics 2 and performed an ANCOVA with gender and grades as independent

TABLE VII. Self-efficacy scores in physics 1 binned by course grade (A, B, C) for female and male students ( $M$  = mean,  $SD$  = standard deviation,  $N$  = number of students), along with Cohen’s  $d$  for the gender contrast in means.

Post self-efficacy	Physics 1								
	C			B			A		
	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$
Female	2.33	0.42	86	2.59	0.45	80	2.90	0.41	31
Male	2.75	0.42	134	2.90	0.43	173	3.14	0.38	105
$d$ size	0.89 <sup>b</sup>	...	...	0.67 <sup>b</sup>	...	...	0.60 <sup>a</sup>	...	...

$d$  sizes are Cohen’s  $d$  values between female and male students’ self-efficacy scores.

<sup>a</sup>Significant at the 0.01 probability level.

<sup>b</sup>Significant at the 0.001 probability level.

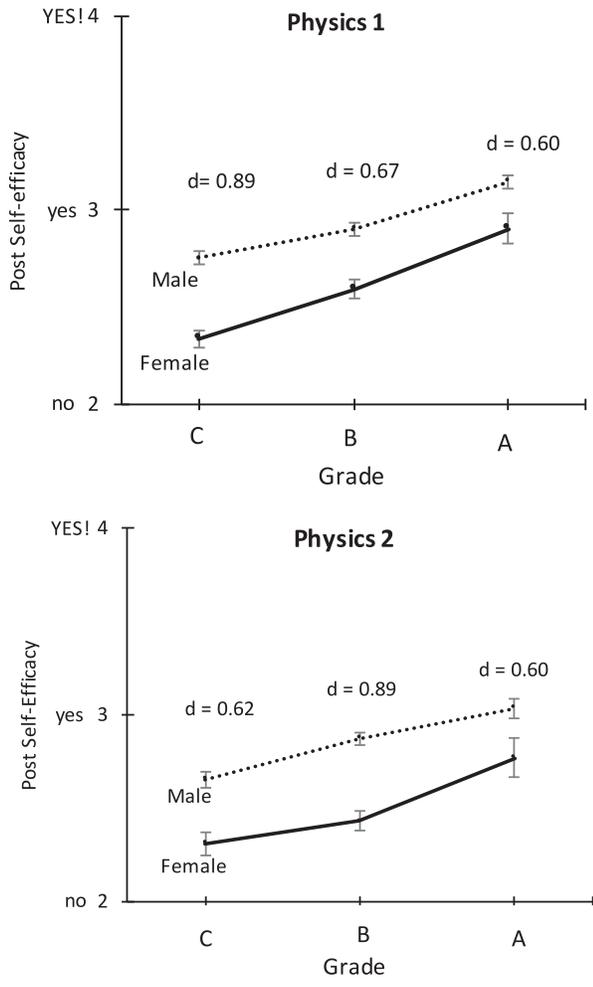


FIG. 2. Post-self-efficacy scores of female and male students binned by course grade (A, B, C) in physics 1 (top) and physics 2 (bottom) courses. Error bars represent standard error of the mean.

variables and post self-efficacy as dependent variable. We found that the mean value of the gender gap in self-efficacy decreased only a small amount to 0.34 ( $p < 0.001$ ) in physics 1 and 0.38 ( $p < 0.001$ ) in physics 2 courses.

In other words, a small part of the gender gap in self-efficacy might be attributable to differences in performance, but the gap mainly comes from biased perceptions: 87% in physics 1 and 95% in physics 2 (see Fig. 3). Consistent with our results discussed earlier, existing gender gap in students' self-efficacy is largely due to students' self-perception rather than how they actually perform in the course. In other words, gender gap in self-efficacy does not come from performance differences by gender but mainly from biased beliefs about physics among female and male students.

We also repeated the analysis with each self-efficacy survey question. This analysis allowed us to verify that the gender difference in self-efficacy was not dominated by only a few aspects of self-efficacy such as test-taking or in-class performance. Specifically, we performed a two-way ANOVA in which the dependent variable is the student's score on the survey item and the independent variables are gender (coded as 0 for female, 1 for male students) and course grade (0 to 4) in physics 1 or physics 2. The standardized regression coefficients for gender, denoted by  $\beta_1$  in Eq. (1) are shown in Table IX. There was always a statistically significant gender effect—male students were more positive in their responses to all of the self-efficacy questions than female students in both physics 1 and physics 2 courses, even when taking into account students' course grades. The effects were slightly smaller for the questions about lab classes and physics research, but generally were consistent in size across questions.

Self-efficacy survey item score

$$= \beta_1 \times \text{gender} + \beta_2 \times \text{grades} + \text{const.} \quad (1)$$

**C. Gender differences in self-efficacy across different instructors and course types**

We also explored whether gender differences in self-efficacy exist for different instructors and teaching methods in introductory level physics courses to determine the generalizability of the findings of research questions

TABLE VIII. Self-efficacy scores in physics 2 binned by course grade (A, B, C) for female and male students ( $M$  = mean,  $SD$  = standard deviation,  $N$  = number of students), along with Cohen's  $d$  for the gender contrast in means.

Post self-efficacy	Physics 2								
	C			B			A		
	$M$	$SD$	$N$	$M$	$SD$	$N$	$M$	$SD$	$N$
Female	2.30	0.51	70	2.43	0.49	87	2.76	0.54	27
Male	2.65	0.53	133	2.87	0.43	158	3.03	0.47	81
$d$ size	0.62 <sup>b</sup>	...	...	0.89 <sup>b</sup>	...	...	0.60 <sup>a</sup>	...	...

$d$  sizes are Cohen's  $d$  values between female and male students' self-efficacy scores.

<sup>a</sup>Significant at the 0.05 probability level.

<sup>b</sup>Significant at the 0.001 probability level.

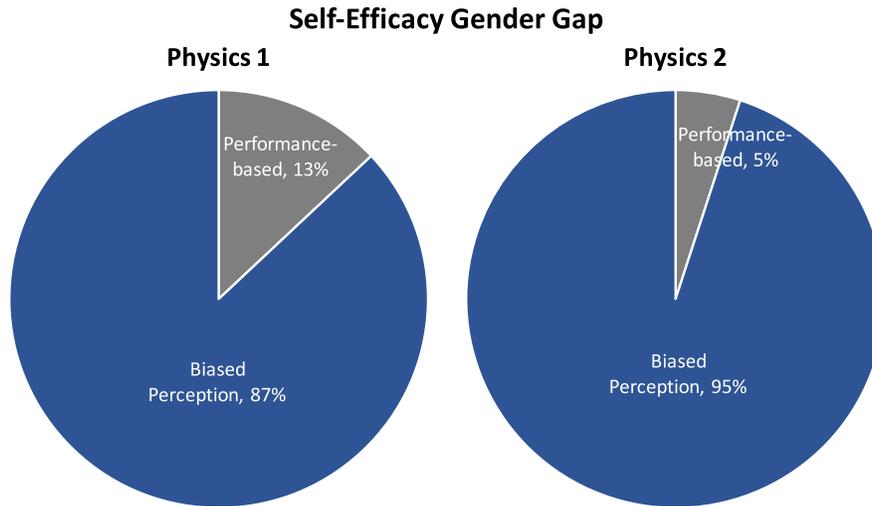


FIG. 3. The relative contributions of performance and biased perception to the post-self-efficacy gender gap in each course.

1 and 2. There were 9 sections in physics 1 and 11 sections of physics 2 courses in the data set. All of the instructors were male, but with some having several years of teaching experience and others new to teaching. Some of these sections were led by the same instructor, and thus not a meaningful generality test, leading us to merge sections within physics 1 or within physics 2 that were taught by the same instructor. However, we did not merge sections across courses or if they were of different formats within a course (*traditional vs flipped*). As a result, we analyzed four sections of physics 1 (three traditional sections and 1 flipped section) and four sections of physics 2 (three traditional sections and 1 flipped section).

To formally quantify the gender effect on self-efficacy controlling for performance across the various physics courses, we performed a linear regression within each section in which post self-efficacy is the dependent variable, and gender and physics performance level (FCI or CSEM as appropriate) are the independent variables [see Eqs. (2) and (3)]. Of particular interest is the  $\beta_1$  estimate (the gender effect) within each section.

Post self-efficacy in physics 1

$$= \beta_1 \times \text{gender} + \beta_2 \times \text{FCI prescores} + \text{const}, \quad (2)$$

Post self-efficacy in physics 2

$$= \beta_1 \times \text{gender} + \beta_2 \times \text{CSEM prescores} + \text{const}. \quad (3)$$

As shown in Table X, the gender effects were statistically significant in all but one case across courses, instructors, and formats, and that one exception case had a similar sized  $\beta_1$  and a small number of students, so the most likely explanation is low statistical power. Figure 4 plots the  $\beta_1$  values across sections: it is clear that the gap in self-efficacy for female and male students did not diminish in flipped, i.e., active engagement courses in either physics 1 or physics 2. For example, in physics 1, the standardized beta coefficient for gender  $\beta_1$  in the flipped course ( $\sim 0.3$ ) was similar to the  $\beta_1$  in the traditional courses (between  $\sim 0.2 - 0.4$ ). Similarly, in physics 2, the standardized beta coefficient for gender  $\beta_1$

TABLE IX. Gender effect  $\beta_1$  values for each self-efficacy question at pretest and post-test in physics 1 and physics 2 courses. *R* indicates that the survey item was reverse coded.

Survey Item	Physics 1		Physics 2	
	Pre	Post	Pre	Post
1. I can complete the physics activities I get in a lab class	0.14 <sup>a</sup>	0.21 <sup>b</sup>	0.19 <sup>c</sup>	0.19 <sup>a</sup>
2. If I went to a museum, I could figure out what is being shown about physics in:	0.22 <sup>c</sup>	0.27 <sup>c</sup>	0.24 <sup>a</sup>	0.28 <sup>c</sup>
3. I am often able to help my classmates with physics in the laboratory or in recitation.	0.19 <sup>c</sup>	0.23 <sup>c</sup>	0.25 <sup>c</sup>	0.25 <sup>c</sup>
4. I get a sinking feeling when I think of trying to tackle difficult physics problems. (R)	0.24 <sup>c</sup>	0.24 <sup>c</sup>	0.25 <sup>c</sup>	0.27 <sup>c</sup>
5. If I wanted to, I could be good at doing physics research.	0.17 <sup>a</sup>	0.17 <sup>c</sup>	0.17 <sup>a</sup>	0.18 <sup>b</sup>
6. If I study, I will do well on a physics test.	0.18 <sup>b</sup>	0.26 <sup>c</sup>	0.24 <sup>c</sup>	0.26 <sup>c</sup>

<sup>a</sup>Significant at the 0.05 probability level.

<sup>b</sup>Significant at the 0.01 probability level.

<sup>c</sup>Significant at the 0.001 probability level.

TABLE X. For different instructors and class types in physics 1 and physics 2 courses, number of students ( $N$ ), gender standardized coefficient ( $\beta_1$ ), and  $p$  values are given.

Class number	Section, instructor, class type	$N$	$\beta_1$ values	$p$ values
Physics 1	Instructor 1 Traditional	140	0.19	0.018 <sup>a</sup>
	Instructor 2 Traditional	26	0.18	0.95 <sup>c</sup>
	Instructor 3 Traditional	25	0.42	0.017 <sup>a</sup>
	Instructor 4 Flipped	128	0.30	0.001 <sup>a</sup>
Physics 2	Instructor 4 Traditional	112	0.34	<0.001 <sup>b</sup>
	Instructor 4 Flipped	286	0.29	<0.001 <sup>b</sup>
	Instructor 5 Traditional	96	0.35	<0.001 <sup>b</sup>
	Instructor 6 Traditional	101	0.29	0.003 <sup>a</sup>

<sup>a</sup>Significant at the 0.01 probability level.

<sup>b</sup>Significant at the 0.001 probability level.

<sup>c</sup>Nonsignificant result.

in the flipped course ( $\sim 0.3$ ) is similar to the  $\beta_1$  in the traditional courses (between  $\sim 0.3 - 0.4$ ). Even though flipped instruction might improve learning and possibly self-efficacy overall, it does not decrease this gender effect.

#### IV. SUMMARY AND DISCUSSION

Extensive prior research has shown that self-efficacy in science is an important driver of interest in science [40], STEM career choice [69], and persistence towards those career goals [48]. Prior research has also shown that female

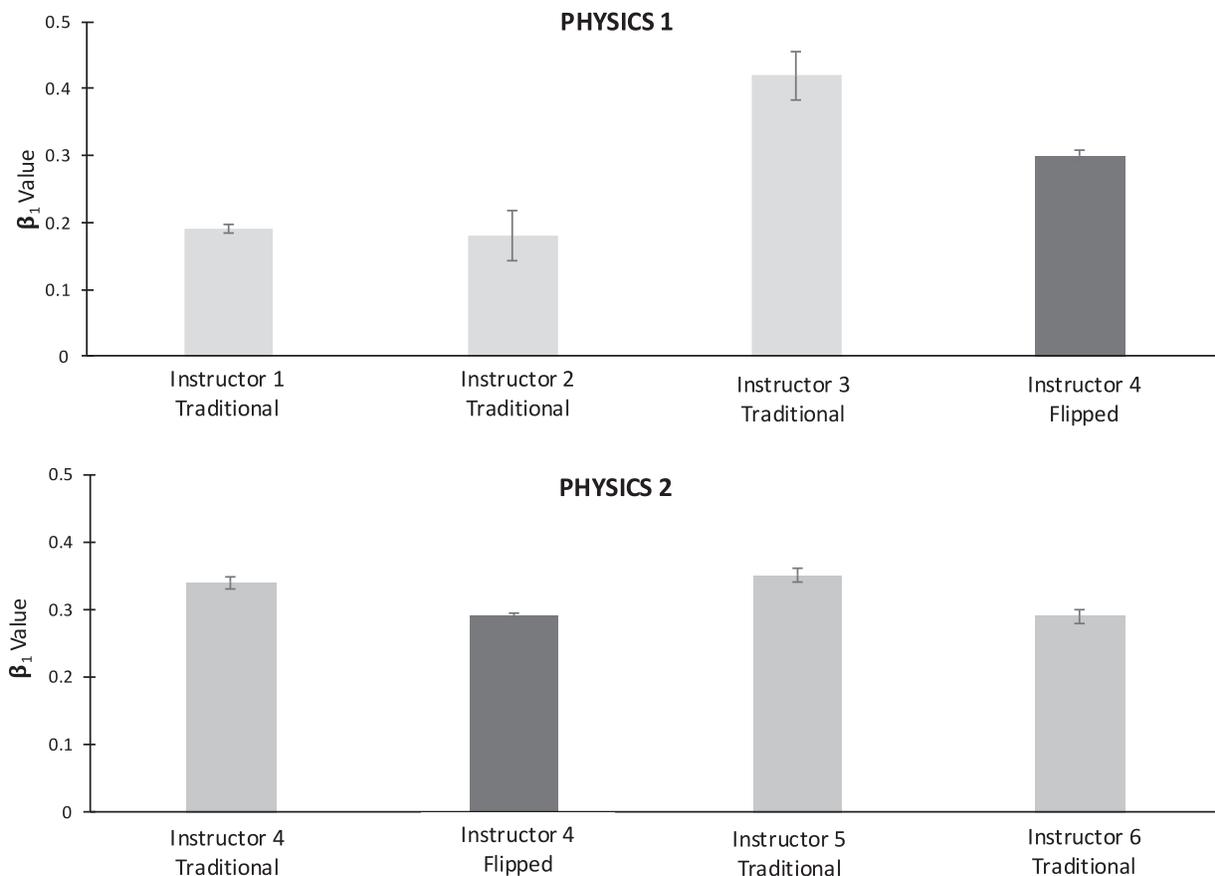


FIG. 4. Gender standardized coefficients ( $\beta_1$  values) for different instructors and class types in physics 1 (top row) and physics 2 (bottom row) courses. Darker and lighter bars represent flipped and traditional physics courses, respectively. Error bars represent standard error of the  $\beta_1$  values.

students have significantly lower self-efficacy than male students in STEM fields [23,53], and this gap in self-efficacy partly contributes to the underrepresentation of women in science and engineering fields [68,69]. However, few studies have focused on self-efficacy differences between female and male students controlling for performance, either to remove a natural confound or to understand how the self-efficacy gap might vary across performance levels. Therefore, we investigated self-efficacy differences of male and female students with equal performance levels in introductory physics courses, which are generally recognized as gateway courses for obtaining a physical science or engineering degree.

Our findings indicate that female students had significantly lower self-efficacy than male students throughout a two-semester introductory physics course sequence at every matched performance level. Further, the gap persists regardless of whether performance was measured through research-validated instruments or through the performance indicators provided to students (i.e., grades). Importantly, the gender self-efficacy gap grew after instruction in physics 1 and physics 2 courses, most notably within the highest achieving student group, in both traditional and flipped courses. The findings suggest that the self-efficacy of female students, and especially high achieving female students, is negatively impacted by their experiences in introductory physics courses.

Furthermore, the gender gap in students' post self-efficacy is found to largely stem from students' perceptions in the course and the contribution from students' actual performance is very small. There can be several possible reasons for students' incorrect judgements about their competence. Most saliently, societal stereotypes and cultural beliefs about gender and STEM achievement can bias students' self-assessment of their competence. For instance, female and male students' self-confidence is likely to be influenced by beliefs about the discipline and who can succeed in it. Further, as noted in the background section, women can be subject to implicit or explicit stereotype threat in physics courses [97,98]. Negative stereotypes about women in physics may cause women to have a different perception of success than men when they initially enter in a male-dominated discipline in which contributions of "brilliant" men are overemphasized. Female students may even assume that they have to make extra efforts to succeed in physics relative to male students and their achievement is not a reflection of how good they are in physics unlike the achievements of "successful" men. Likewise, women might undergo additional stress and struggle to demonstrate their skills to be valued equally as men in a classroom in which they are underrepresented. Also, as mentioned in the introduction, some prior research has found that teaching practices and interactions may treat female students differently than male students. For example, if female students are not called upon to answer questions or not given the same type of positive feedback

as male students, this could have a negative effect on self-efficacy. In addition, students generally compare themselves to others who are similar where gender can be one determinant of being similar [24]. The thought of "there are not many people like me" can negatively influence women's self-efficacy and reinforce stereotypical beliefs about women's ability in physics. Similarly, men may portray higher confidence in their ability regardless of how they perform due to these biased perceptions which favor their gender. Correll found that men assess their math competence higher than women even though they perform similarly [68]. In this research, Correll also found that boys were more likely to pursue careers requiring math competence and skills at higher rates than girls, not because they were actually better at math but rather because they *thought* they were better. Describing her research on gender differences in math, Tobias notes that "when girls succeed at a math lesson or on a math quiz, they attribute their success to luck; boys attribute it to their own inner ability. When girls fail, they attribute their failure to a lack of ability; boys attribute theirs to a lack of effort. That's why even girls who do well in mathematics in school do not develop the kind of confidence males do." [99]. This type of dichotomy in how male and female students internalize their successes and failures may be partly responsible for the lower self-efficacy of women in physics courses. In summary, these types of environmental and sociocultural biases and gender-based beliefs about physics can have a large impact on students' self-beliefs in their competence than how they actually perform in the course. Moreover, it is possible that if female students had higher self-efficacy about physics, they would have less anxiety about learning physics and all of their cognitive resources while solving physics problems would be devoted to learning, which would have the potential to boost their performance to a higher level than what we observed in this study.

### A. Implications for the physical sciences and engineering

The observed effect is particularly alarming because of its large size. In a research sense, the effect can be considered large because the Cohen's  $d$  approached 1. In a more practical sense, the effect can be considered large because we found that female students with high scores on physics conceptual surveys (or who are receiving A's) had similar self-efficacy as male students with only medium or low scores on physics conceptual surveys (or are receiving B's and C's). Whether this effect is framed as underconfidence among women, overconfidence among men, or some combination (perhaps the most likely interpretation), the practical outcome could be the same: worse outcomes for women in STEM.

In the introductory physics course sequence, the self-efficacy gaps may produce higher levels of anxiety during exams, negatively impacting exam performance [100]. In addition, self-efficacy problems have been shown to impact

interest, and therefore high achieving female students may begin to lose interest in engineering and physical science due to their inaccurate assessment of their capability in physics. Self-efficacy is also related to self-regulated learning strategies, that is, higher self-efficacy is associated with better self-regulated learning [44]. This decreased interest and the lower self-efficacy may trigger female students to devote less time to homework or studying, or even to drop out of physics courses or decide to exit the STEM major track altogether.

In sum, inaccurate assessments of one's capability and/or performance can influence interests and career goals. Gender differences in self-efficacy, especially in fields which have been historically male dominated, may inhibit progress toward increasing the diversity in these fields. We discovered alarming trends in female students' self-efficacy in introductory physics courses which are generally required for engineering and other STEM majors—in particular, female students with A grades often had similar physics self-efficacy as male students with C grades. These self-efficacy gaps may result in an “accumulation of disadvantage” for women in physical science and engineering domains. In other words, female students' lower self-efficacy in physics than male students may partly contribute to the underrepresentation of women (and even highly qualified women) in some STEM fields. Even minor gender differences in self-efficacy in early college experiences can add up to major inequalities later in STEM careers [101,102].

The current research examined only physics, but similar patterns are likely to occur in other courses for which there are negative gender stereotypes such as difficult mathematics, engineering, computer science, and other STEM courses. For example, research has already shown that male students have higher self-efficacy than female students of similar performance in mathematics [68]. Such broader gender gaps in STEM-related self-efficacies may result in differential educational experiences between male and female students who major in STEM-related fields and, ultimately, the underrepresentation of women in those fields. Thus, it is imperative to reflect on ways to broadly improve the environments of introductory STEM courses for which there are negative gender stereotypes in order to help female students reconcile their self-efficacy with their actual performance and capability.

### **B. Implications for instruction**

What might be causing the differences in male and female students' self-efficacy, or alternatively, what might be done to reduce these differences? Since the effect grows with instruction, features of instruction seem particularly important to consider, both the instructional style of the instructor and the general pedagogy used in the class. Past

research suggests that some instructors may have implicit (or even explicit) negative gender biases [75]. Although instructional style was not formally measured, the study included a wide range of instructors (tenure-stream and nontenure stream; some having won teaching awards) and pedagogies (from very traditional lecture to including a number of active learning strategies as part of flipped instruction). Yet we found that the gender differences in self-efficacy were consistent across the instructors and class types, i.e., lecture-based and active-engagement courses. Thus, whatever the source, the recently explored adjustments to pedagogy to increase learning appear not to be relevant to this gender gap in self-efficacy [103,104]. It may be that messages conveyed by students to each other [25] or the broader culture [55] may at least partly be the root cause and thus a different kind of countermessaging is required.

Why might active learning not reduce the gender gaps in self-efficacy? While active learning may be beneficial for both male and female students in terms of performance outcomes, the nature of these in-class interactions may result in a decrease in female students' self-efficacy. Felder *et al.* [25] note that women usually play less active roles than men in cooperative learning groups in engineering and instead women report feeling that group work benefitted them because there were opportunities to have the material explained to them (i.e., reinforcing stereotypes of relative weakness). In addition, women also report feeling that their contributions in group work are undervalued and their contributions in active learning situations may also be ignored or discounted by other male students in the group [25]. Therefore, it is important for instructors to think carefully about how active learning is implemented and ways to help all students benefit from it. For example, instructors may need to frequently remind students that all group members' contributions are important and valuable. In addition, cooperative learning groups can be structured such that women outnumber men in any group containing women—in this way, women may not feel as intimidated by male members in the group. Investigations of the types of pedagogies and interventions that help women accurately assess their capability and performance in introductory physics courses are crucial in reducing the alarming self-efficacy gender gaps and their detrimental effects.

### **ACKNOWLEDGMENTS**

This work was supported by Grant No. DUE-1524575 from the National Science Foundation and Grant No. 220020483 from the James S. McDonnell Foundation. Portions of these results were presented at the 2018 American Association of Physics Teachers Summer Meeting and Physics Education Research Conference.

- [1] National Student Clearinghouse Research Center (2015). Science and Engineering Degree Attainment: 2004–2014, <https://nscresearchcenter.org/snapshotreport-degree-attainment15/>.
- [2] <https://www.aip.org/statistics/data-graphics/percent-physics-bachelors-and-phds-earned-women-classes-1975-through-2016>; Women, Minorities, and Persons with Disabilities in Science and Engineering, Occupation (2017), <https://nsf.gov/statistics/2017/nsf17310/digest/occupation/women.cfm>.
- [3] Degrees in engineering and engineering technologies conferred by postsecondary institutions, by level of degree and sex of student: Selected years, 1949–50 through 2013–14 (2015), [https://nces.ed.gov/programs/digest/d15/tables/dt15\\_325.45.asp](https://nces.ed.gov/programs/digest/d15/tables/dt15_325.45.asp).
- [4] B. L. Yoder, Engineering by the numbers Washington, DC: ASEE, <https://www.asee.org/papers-and-publications/publications/college-profiles/15EngineeringbytheNumbersPart1.pdf>.
- [5] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, Exploring the gender gap in the conceptual survey of electricity and magnetism, *Phys. Rev. Phys. Educ. Res.* **13**, 020114 (2017).
- [6] A. Traxler, R. Henderson, J. Stewart, G. Stewart, A. Papak, and R. Lindell, Gender fairness within the Force Concept Inventory, *Phys. Rev. Phys. Educ. Res.* **14**, 010103 (2018).
- [7] A. Traxler, X. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, *Phys. Rev. Phys. Educ. Res.* **12**, 020114 (2016).
- [8] A. Traxler and E. Brewe, Equity investigation of attitudinal shifts in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020132 (2015).
- [9] Z. Hazari, G. Potvin, R. Lock, F. Lung, G. Sonnert, and P. Sadler, Factors that affect the physical science career interest of female students: Testing five common hypotheses, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020115 (2013).
- [10] Z. Hazari, G. Sonnert, P. Sadler, and M. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, *J. Res. Sci. Teach.* **47**, 978 (2010).
- [11] Z. Hazari, P. Sadler, and G. Sonnert, The science identity of college students: Exploring the intersection of gender, race, and ethnicity, *J. Coll. Sci. Teach.* **42**, 82 (2013).
- [12] R. Ivie, S. White, and R. Chu, Women’s and men’s career choices in astronomy and astrophysics, *Phys. Rev. Phys. Educ. Res.* **12**, 020109 (2016).
- [13] R. Ivie and K. Stowe, Women in Physics, 2000, <https://eric.ed.gov/?id=ED446559>.
- [14] M. Besterfield-Sacre, M. Moreno, L. Shuman, and C. Atman, Gender and ethnicity differences in freshmen engineering student attitudes: A cross-institutional study, *J. Eng. Educ.* **90**, 477 (2001).
- [15] C. Seron, S. Silbey, S. Cech, and B. Rubineau, Persistence is cultural: Professional socialization and the reproduction of sex segregation, *Work and Occupations* **43**, 178 (2016).
- [16] R. Felder, G. Felder, and E. Dietz, Longitudinal study of engineering student performance and retention. V. Comparisons with traditionally-taught students, *J. Eng. Educ.* **87**, 469 (1998).
- [17] A. W. Astin, *What matters in college: Four critical years revisited* (Jossey-Bass, San Francisco, CA, 1993).
- [18] N. M. Hewitt and E. Seymour, A long, discouraging climb, *ASEE Prism* (1992), Vol. 1, p. 24.
- [19] E. Seymour and N. M. Hewitt, *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997).
- [20] S. Tobias, They’re Not Dumb, They’re Different: Stalking the Second Tier, Tuscon, AZ: Research Corporation (1990).
- [21] S. Cheryan, S. Ziegler, A. Montoya, and L. Jian, Why are some STEM fields more gender balanced than others?, *Psychol. Bull.* **143**, 1 (2017).
- [22] M. Wang and J. Degol, Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields, *Developmental Review* **33**, 304 (2013).
- [23] S. Cheryan and V. Plaut, Explaining underrepresentation: A theory of precluded interest, *Sex Roles* **63**, 475 (2010).
- [24] C. Hill, C. Corbett, and A. Rose, Why so few? Women in science, technology, engineering, and mathematics (2010), <https://www.aauw.org/files/2013/02/Why-So-Few-Women-in-Science-Technology-Engineering-and-Mathematics.pdf>. <http://onlinelibrary.wiley.com/doi/10.1002/j.2168-9830.2010.tb01066.x/epdf>.
- [25] R. Felder, G. Felder, M. Mauney, C. Hamrin, and E. Dietz, A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes, *J. Eng. Educ.* **84**, 151 (1995).
- [26] E. Lichtenberger and C. George-Jackson, Predicting high school students’ interest in majoring in a STEM field: Insight into high school students’ postsecondary plans, *J. Career Techn. Educ.* **28**, 19 (2013).
- [27] C. Nord *et al.*, The nation’s report card: America’s high school graduates (NCES 2011-462). Washington, CD: U.S. Government Printing Office (2011), <https://files.eric.ed.gov/fulltext/ED518324.pdf>.
- [28] A. Strenta, R. Elliot, R. Adair, M. Matier, and J. Scott, Choosing and leaving science in highly selective institutions, *Res. High. Educ.* **35**, 513 (1994).
- [29] E. Pascarella and P. Terenzini, *How College Affects Students: Findings and Insights from Twenty Years of Research* (Jossey-Bass, San Francisco, CA, 1991).
- [30] A. Wigfield *et al.* Motivation to succeed, in *Handbook of Child Psychology, Social Emotional and Personality Development*, edited by W. Damon and N. Eisenberg (Wiley New York, 2006), Vol. 3.
- [31] M. Lorenzo, C. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).
- [32] R. Beichner *et al.*, The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) Project (2007), [http://www.per-central.org/per\\_reviews/volume1.cfm?](http://www.per-central.org/per_reviews/volume1.cfm?).
- [33] A. Diekman, E. Brown, J. Johnston, and E. Clark, Seeking congruity between goals and roles: A new look at why women opt out of STEM careers, *Psychol. Sci.* **21**, 1051 (2010).

- [34] J. Smith, C. Sansone, and P. White, The stereotyped task engagement process: The role of interest and achievement motivation. *J. Educ. Psychol.* **99**, 99 (2007).
- [35] S. Ahlqvist, B. London, and L. Rosenthal, Unstable identity compatibility: How gender rejection sensitivity undermines the success of women in science, technology, engineering and mathematics fields, *Psychol. Sci.* **24**, 1644 (2013).
- [36] C. Dweck, Is math a gift? Beliefs that put females at risk, *Why Aren't More Women In Science? Top Researchers Debate the Evidence*, edited by S. Ceci and W. Williams (APA Press Washington, DC, 2007), pp. 47–55.
- [37] C. Dweck, Mindsets and math/science achievement, institute for advanced study, *Commission on Mathematics and Science Education* (Carnegie Corporation of New York, New York, 2008).
- [38] I. Goodman, C. Cunningham, and C. Lachapelle, The Women's Experience in College Engineering (2002), <https://files.eric.ed.gov/fulltext/ED507395.pdf>.
- [39] R. Lapan, A. Adams, S. Turner, and J. Hinkelman, Seventh graders' vocational interest and efficacy expectation patterns, *J. Career Develop.* **26**, 215 (2000).
- [40] J. Eccles, Understanding women's educational, and occupational choices: Applying the Eccles, *et al.* model of achievement-related choices, *Psychol. Women Q.* **18**, 585 (1994).
- [41] A. Bandura, Self-efficacy, in *Encyclopedia of Psychology*, 2nd ed., edited by R. J. Corsini (Wiley, New York 1994), Vol. 3, pp. 368–369.
- [42] A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1997).
- [43] F. Pajares, Gender differences in mathematics self-efficacy beliefs, *Gender Differences in Mathematics: An Integrative Psychology Approach* (Cambridge University Press, New York, 2005), pp. 294–315.
- [44] B. Zimmerman, Self-efficacy: An essential motive to learn, *Contemp. Educ. Psychol.* **25**, 82 (2000).
- [45] D. Schunk and F. Pajares, The development of academic self-efficacy, *Development of Achievement Motivation: A Volume in the Educational Psychology Series*, edited by A. Wigfield and J. Eccles (Academic Press, San Diego, CA, 2002), pp. 15–31.
- [46] H. Watt, The role of motivation in gendered educational and occupational trajectories related to maths, *Educ. Res. Eval.* **12**, 305 (2006).
- [47] A. Zeldin, S. Britner, and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *J. Res. Sci. Teach.* **45**, 1036 (2008).
- [48] N. Betz and G. Hackett, Applications of self-efficacy theory to understanding career choice behavior, *J. Soc. Clin. Psychol.* **4**, 279 (1986).
- [49] S. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, *J. Res. Sci. Teach.* **43**, 485 (2006); S. L. Britner, Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes, *J. Res. Sci. Teach.* **45**, 955 (2008); A. L. Zeldin, S. L. Britner, and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *J. Res. Sci. Teach.* **45**, 1036 (2008).
- [50] A. Bandura, Social cognitive theory of self-regulation, *Organ. Behav. Hum. Decis. Process.* **50**, 248 (1991).
- [51] P. Pintrich, A motivational science perspective on the role of student motivation in learning and teaching contexts, *J. Educ. Psychol.* **95**, 667 (2003).
- [52] T. Bouffard-Bouchard, S. Parent, and S. Larivee, Influence of self-efficacy on self-regulation and performance among junior and senior high-school aged students, *Int. J. Behav. Dev.* **14**, 153 (1991).
- [53] P. Huang and S. Brainard, Identifying determinants of academic self-confidence among science, math, engineering, and technology students, *J. Women Minorities Sci. Engin.* **7**, 315 (2001).
- [54] P. Vincent-Ruz and C. Schunn, The increasingly important role of science competency beliefs for science learning in girls, *J. Res. Sci. Teach.* **54**, 790 (2017).
- [55] S. Brainard and L. Carlin, A six-year longitudinal study of undergraduate women in engineering and science, *J. Eng. Educ.* **87**, 369 (1998).
- [56] L. Irani, Understanding gender and confidence in CS course culture, *SIGCSE 2004, Norfolk VA* (2004), <http://ai2-s2-pdfs.s3.amazonaws.com/f5c2/cd1620808de342808613cd0cd4a67dbd8931.pdf>.
- [57] M. Rosson, J. Carroll, and H. Sinha, Orientation of undergraduates toward careers in the computer and information sciences: Gender, self-efficacy and social support, *ACM Trans. Comput. Educ.* **11**, 1 (2011).
- [58] A. Cavallo, W. Potter, and M. Rozman, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors, *School Sci. Math.* **104**, 288 (2004).
- [59] V. Sawtelle, E. Brewe, and L. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *J. Res. Sci. Teach.* **49**, 1096 (2012).
- [60] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020105 (2016).
- [61] C. Lindstrøm and M. D. Sharma, Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?, *Int. J. Innov. Sci. Math. Educ.* **19**, 1 (2011).
- [62] E. Marshman, Z. Y. Kalender, C. Schunn, T. Nokes-Malach, and C. Singh, A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences, *Can. J. Phys.* **96**, 391 (2018).
- [63] R. Marra, K. Rodgers, D. Shen, and B. Bogue, Women engineering students and self-efficacy: A multi-year, multi-institution study of women engineering student self-efficacy, *J. Eng. Educ.* **98**, 27 (2009).
- [64] J. Raelin, M. Bailey, J. Hamann, L. Pendleton, R. Reisberg, and D. Whitman, The gendered effect of cooperative education, contextual support, and self-efficacy on undergraduate retention, *J. Eng. Educ.* **103**, 599 (2014).
- [65] O. Eris, D. Chachra, H. L. Chen, S. Sheppard, L. Ludlow, C. Rosca, T. Bailey, and G. Toye, Outcomes of a

- longitudinal administration of the persistence in engineering survey, *J. Eng. Educ.* **99**, 371 (2010).
- [66] K. Meyers, S. Silliman, N. Gedde, and M. Ohland, A comparison of engineering students' reflections on their first-year experiences, *J. Eng. Educ.* **99**, 169 (2010).
- [67] B. M. Jones, Pursuing diversity at state-supported residential STEM schools, *NCSSMST Journal* **16**, 30 (2010).
- [68] S. Correll, Gender and the career choice process: The role of biased self-assessments, *Am. J. Sociology* **106**, 1691 (2001).
- [69] S. Correll, Constraints into preferences: Gender, status, and emerging career aspirations, *Am. Sociological Rev.* **69**, 93 (2004).
- [70] M. Niederle and L. Vesterlund, Do women shy away from competition? Do men compete too much?, *Q. J. Econ.* **122**, 1067 (2007).
- [71] P. Rosser, The SAT Gender Gap: Identifying the Causes (1989), <https://eric.ed.gov/?id=ED311087>.
- [72] R. Childs, Gender Bias and Fairness (1990), <http://pareonline.net/getvn.asp?v=2andn=3>.
- [73] M. Spear, Sex bias in science teachers' ratings of work and pupil characteristics, *Eur. J. Sci. Educ.* **6**, 369 (1984).
- [74] D. Sadker, M. Sadker, and K. Zittleman, *Still Failing at Fairness: How Gender Bias Cheats Girls and Boys in School* (Scribner, New York, 1994).
- [75] C. Moss-Racusin, J. Dovidio, V. Brescoll, M. Graham, and J. Handelsman, Science faculty's subtle gender biases favor male students, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474 (2012).
- [76] V. Coletta, J. Phillips, and J. Steinert, FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects, *AIP Conf. Proc.* **1413**, 23 (2012).
- [77] L. McCullough, Gender, context, and physics assessment, *J. Int. Wom. Stud.* **5**, 20 (2004).
- [78] A. Wigfield, J. Eccles, and P. Pintrich, Development between the ages of eleven and twenty-five, in *The Handbook of Educational Psychology*, edited by D. Berliner and R. Calfee (MacMillan Publishing, New York, 1996).
- [79] J. Blickenstaff, Women and science careers: Leaky pipeline or gender filter, *Gender Educ.* **17**, 369 (2005).
- [80] J. Enfield, Looking at the impact of the flipped classroom model of instruction on Undergraduate Multimedia Students at CSUN, *TechTrends* **57**, 14 (2013).
- [81] L. P. Galway, K. K. Corbett, T. K. Takaro, K. Tairyan, and E. Frank, A novel integration of online and flipped classroom instructional models in public health higher education, *BMC Med. Educ.* **14**, 181 (2014).
- [82] H. Fencil and K. Scheel, Engaging students: An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course, *J. Coll. Sci. Teach.* **35**, 20 (2005).
- [83] E. Brewster, Modeling theory applied: Modeling Instruction in introductory physics, *Am. J. Phys.* **76**, 1155 (2008).
- [84] V. Sawtelle, E. Brewster, and L. Kramer, Positive impacts of Modeling Instruction on self-efficacy, *AIP Conf. Proc.*, **1289**, 289 (2010).
- [85] G. Novak, E. Patterson, A. Gavrin, and W. Christian, *Just-In-Time Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999).
- [86] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [87] Activation Lab Science Technology Engineering Art Mathematics (2017), <http://www.activationlab.org/tools/>; C. Dweck, *Mindset: The New Psychology of Success* (Ballentine, New York, 2006); *Self-theories: Their Role in Motivation, Personality, and Development* (Psychology Press, Philadelphia, 1999).
- [88] J. Schell and B. Lukoff, Peer instruction self-efficacy instrument [Developed at Harvard University] (unpublished); E. Redish, J. Saul, and R. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- [89] S. Glynn, P. Brickman, N. Armstrong, and G. Taasoobshirazi, Science motivation questionnaire II: Validation with science majors and nonscience majors, *J. Res. Sci. Teach.* **48**, 1159 (2011).
- [90] Z. Y. Kalender, E. Marshman, T. Nokes-Malach, C. Schunn, and C. Singh, Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses, in *Proceedings of the Physics Education Research Conference 2017, Cincinnati, OH* (AIP, New York, 2018), pp. 204–207, DOI:10.1119/perc.2017.pr.046.
- [91] T. Nokes-Malach, E. Marshman, Z. Y. Kalender, C. Schunn, and C. Singh, Investigation of male and female students' motivational characteristics throughout an introductory physics course sequence, in *Proceedings of the Physics Education Research Conference 2017, Cincinnati, OH* (AIP, New York, 2018), pp. 276–279, DOI:10.1119/perc.2017.pr.064.
- [92] L. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* **16**, 297 (1951).
- [93] J. Nunnally, *Psychometric Theory*, 2nd ed. (McGraw-Hill, New York, 1978); M. Chi, Thinking aloud, in *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes*, edited by M. W. Van Someren, Y. F. Barnard, and J. A. C. Sandberg (Academic Press, London, 1994).
- [94] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [95] C. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- [96] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, Hillsdale, NJ, 1988).
- [97] A. Maries, N. Karim, and C. Singh, Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?, *Phys. Rev. Phys. Educ. Res.* **14**, 020119 (2018).
- [98] G. C. Marchand and G. Taasoobshirazi, Stereotype threat and women's performance in physics, *Int. J. Sci. Educ.* **35**, 3050 (2013).
- [99] S. Tobias, Rethinking teaching math, science, *Education World*, 2002; *Banishing Math Anxiety* (Kendall Hunt, Dubuque, IA, 2012).
- [100] P. Pintrich and E. De Groot, Motivational and self-regulated learning components of classroom academic performance, *J. Educ. Psychol.* **82**, 33 (1990).

- [101] L. Babcock and S. Laschever, *Women Don't Ask: Negotiation and the Gender Divide* (Bantam Books, New York, 2007).
- [102] V. Valian, *Why So Slow? The Advancement of Women* (MIT Press, Cambridge, MA, 1997).
- [103] N. Karim, A. Maries, and C. Singh, Impact of evidence-based flipped or active-engagement non-flipped courses on student performance in introductory physics, *Can. J. Phys.* **96**, 411 (2018).
- [104] N. Karim, A. Maries, and C. Singh, Do evidence-based active-engagement courses reduce the gender gap in introductory physics?, *Eur. J. Phys.* **39**, 025701 (2018).