



Using the Social Cognitive Theory Framework to Chart Gender Differences in the Developmental Trajectory of STEM Self-Efficacy in Science and Engineering Students

John Stewart¹ · Rachel Henderson² · Lynnette Michaluk³ · Jessica Deshler⁴ · Edgar Fuller⁵ · Karen Rambo-Hernandez⁶

Published online: 18 August 2020

© Springer Nature B.V. 2020

Abstract

According to Bandura’s social cognitive theory, a student’s self-efficacy influences his or her academic and career decisions, and his or her performance outcomes; as such, a student’s self-efficacy changes with time in response to the student’s experiences. Self-efficacy may also vary by academic domain. Differences in STEM self-efficacy have often been reported between men and women. The purpose of this study is to explore the evolution of domain-specific STEM self-efficacy in students in gateway physics and mathematics courses and how academic feedback influences the evolution of these differences with time. Further, this study explored whether gender differences in self-efficacy are consistent across STEM domains and how these differences change in response to academic feedback. Self-efficacy in multiple academic domains (current mathematics/science class, other STEM classes, and intended profession) was assessed at multiple time points with subscales adapted from the Motivated Strategies for Learning Questionnaire. Linear mixed effects modeling was used to understand how academic feedback provided by test scores influenced changes in self-efficacy. Students in all classes expressed different levels of self-efficacy toward different domains with the lowest self-efficacy toward their current class and the highest toward their intended profession. Only the current math/science class self-efficacy of men and women differed significantly, with women expressing lower self-efficacy. The differences in current class self-efficacy were evident very early in the class before substantive class feedback was received. The evolution of self-efficacy within the class and between classes was the same for men and women.

Keywords Gender · STEM course self-efficacy · Impact of feedback · Introductory math and physics

Introduction

Bandura’s social cognitive theory (SCT) (1977) and Lent, Brown, and Hackett’s social cognitive career theory (SCCT) (1994) are recursive theories in which a person’s self-efficacy

influences performance decisions, and thus their performance outcomes. These decisions and their resulting outcomes then influence future self-efficacy (Bandura 1977, 1997; Lent et al. 1994). Bandura (1994, p. 71) defined self-efficacy as “people’s beliefs about their capabilities to produce designated

✉ John Stewart
jstewart1@mail.wvu.edu

Rachel Henderson
hende473@msu.edu

Lynnette Michaluk
lynnette.michaluk@mail.wvu.edu

Jessica Deshler
jessica.deshler@mail.wvu.edu

Edgar Fuller
efuller@fju.edu

Karen Rambo-Hernandez
rambohernandez@tamu.edu

¹ Department of Physics and Astronomy, West Virginia University, Morgantown, WV, USA

² Department of Physics and Astronomy, Michigan State University, East Lansing, MI, USA

³ West Virginia University Center for Excellence in STEM Education, West Virginia University, Morgantown, WV, USA

⁴ Department of Mathematics, West Virginia University, Morgantown, WV, USA

⁵ Department of Mathematics and STEM Transformation Institute, Florida International University, Miami, FL, USA

⁶ Department of Teaching, Learning, and Culture, Texas A&M University, College Station, TX, USA

levels of performance that exercise influence over events that affect their lives” and refers to personal judgments of one’s capabilities to organize and execute courses of action to attain designated goals. Self-efficacy differs conceptually and psychometrically from other closely related constructs, including outcome expectations, self-concept, and self-esteem (Huang 2013; Zimmerman 2000).

Self-efficacy has been identified as one of several non-cognitive constructs most highly correlated with academic performance (Britner 2008; Richardson et al. 2012), and is important in models of retention and persistence (Larson et al. 2014; Lent et al. 1986, 1987; Sawtelle et al. 2012b). In a review of the literature by Usher and Pajares (2008), many studies reported differences in self-efficacy by gender and race/ethnicity. It is believed that these differences partially explain the underrepresentation of women in Science, Technology, Engineering, and Mathematics (STEM) (Cheryan et al. 2017).

While self-efficacy is one of the most studied constructs for understanding academic performance, additional research is still needed to fully understand Bandura’s model. Relatively few studies have measured self-efficacy at multiple time points, and therefore, the evolution of self-efficacy has not been fully explored. The domain-specific nature of self-efficacy has been explored, but the relations of self-efficacy in multiple domains have been less well studied. Thus, the purpose of this study is to explore the nature of self-efficacy across multiple STEM domains and describe its development, both generally and by gender.

We describe the results of two studies: study 1 and study 2. Study 1 measured the self-efficacy of students in mathematics and physics classes for the 2-year period from fall 2015 to spring 2017 with a self-efficacy survey that was administered mid-semester. We hypothesized that women would have consistently lower self-efficacy across STEM domains. When the results were not consistent across STEM domains, a second study, study 2, was performed which measured self-efficacy in the physics classes from fall 2016 to spring 2017 at two time points: one early in the semester and one mid-semester. The effect of feedback, operationalized as test averages, on student self-efficacy was also investigated.

Results of Prior Studies

Self-efficacy beliefs are thought to develop and change based on information from four sources: performance accomplishment (mastery experience), vicarious learning, social persuasion, and physiological arousal (Bandura 1977, 1986, 1993, 1997; Betz and Hackett 1981). Mastery experiences shape efficacy beliefs through the personal interpretation of one’s performances on specific tasks. In general, successful task completion increases self-efficacy beliefs, while failures lower

such beliefs (Bandura 1997). College STEM classes provide mastery feedback in the form of test scores and course grades; a student’s self-efficacy is expected to evolve because of this feedback.

Some studies have shown that mastery experiences are more important to efficacy beliefs for men than women (Zeldin et al. 2008) while other studies have found the opposite (Britner and Pajares 2006). In a review of the literature of the sources of self-efficacy in schools, Usher and Pajares (2008) reported that correlations between mastery experiences and self-efficacy in school ranged from .29 to .67 (median correlation = .58) and that mastery experiences were the only one of the four sources of self-efficacy that were significantly correlated with self-efficacy in every investigation they had reviewed. Vicarious experiences, or social comparisons, refer to the changes in personal expectations that occur as a result of observing others successfully perform or model similar tasks. Zeldin and Pajares (2000) and Sawtelle et al. (2012a, b) report that vicarious experiences and social persuasion experiences may serve as the primary sources of efficacy beliefs for women. Social persuasion represents feedback, judgments, and support from important others that the individual has the skills needed to succeed. Physiological states, referred to as emotional arousal by Bandura (1977), include enjoyment, anxiety, stress, and fatigue associated with actions, can also influence self-efficacy beliefs either positively or negatively, depending on one’s interpretation of the state.

Self-efficacy has emerged as an important factor in academic motivation, goal setting, performance, and persistence (Richardson et al. 2012). Self-efficacy plays a causal role in the development and employment of academic competencies (Schunk and Pajares 2007) and is necessary for effectively utilizing self-regulation skills to achieve mastery (Bandura 1986, 1993; Schunk and Pajares 2007; Zimmerman 2000). Self-efficacy beliefs interact with self-regulated learning processes to mediate academic achievement outcomes (Zimmerman 2000). Furthermore, “SCCT hypothesizes that general cognitive ability (e.g., as indexed by SAT or ACT scores) and past performance (e.g., high school GPA) both influence college student performance (e.g., GPA) and persistence (e.g., retention) in two ways: directly and indirectly through the mediating paths to student’s self-efficacy beliefs and outcome expectations” (Brown et al. 2008, p. 299). According to Zimmerman (2000), self-efficacy has emerged as a highly effective predictor of students’ motivation and learning, and researchers have successfully verified its discriminant validity and convergent validity in predicting such common motivational outcomes as students’ activity choices, effort, and persistence. Above-average self-efficacy in STEM is positively related to STEM task performance, goal achievement, and persistence in STEM majors and careers (Hutchison et al. 2006; Lent et al. 2008; Marra et al. 2009).

Gender Differences in Domain-Specific Self-Efficacy Gender differences in self-efficacy, the relative strength of the sources of self-efficacy, and how self-efficacy functions in larger models of academic performance and retention have been reported almost from the introduction of the construct (Betz and Hackett 1981). Research has generally shown that women's self-efficacy is lower than men's in STEM domains, despite higher STEM achievement (Kost et al. 2009; Larose et al. 2006; Marshman et al. 2018). Huang's meta-analysis (2013) provides a comprehensive summary of STEM domain-specific results. While no overall gender difference in self-efficacy was found, significant differences were found in mathematics and computer science (Hedge's $g = .18$) with men reporting higher self-efficacy. Hedge's g conventions are similar to Cohen's d ; specifically, .20 is a small effect, .50 is a medium effect, and .80 is a large effect. Most studies find that men report higher levels of mathematics self-efficacy in college; however, not all studies find significant differences between men and women (Hall and Ponton 2005).

A substantial number of studies have investigated self-efficacy in engineering; most report that men have higher self-efficacy toward both success in engineering classes and the profession than women (Besterfield-Sacre et al. 2001; Cech et al. 2011; Hackett et al. 1992; Jagacinski 2013; Vogt et al. 2007). Again, the higher self-efficacy of men toward engineering is not universal with some studies finding no significant difference in self-efficacy (Concannon and Barrow 2009, 2012; Mamaril et al. 2016).

Other research has been performed in the STEM domains of biology, chemistry, and physics. Secondary biology students' self-efficacy toward multiple STEM domains was assessed and showed no difference between boys and girls in biology, but large differences in chemistry, mathematics, and physics; boys reported higher self-efficacy in each (Uitto 2014). College chemistry students demonstrated gender differences in self-efficacy toward specific chemistry tasks (Dalgety and Coll 2006) with women reporting higher self-efficacy on all tasks. Again, the difference was not significant in all studies (Villafañe et al. 2014). In physics courses, Lindstrøm and Sharma (2011) reported a difference in the self-efficacy of men and women in introductory college physics with men expressing higher levels of self-efficacy. Shaw (2004) found significant self-efficacy differences for non-STEM major students in a non-mathematical conceptual physics class with men expressing higher self-efficacy, but no significant difference in the calculus-based physics course for STEM majors. In a study investigating high school students, Nissen (2019) found significant differences in self-efficacy between boys and girls in physics classes but no differences in other science and math classes. Cavallo et al. (2004) found that men had significantly higher self-efficacy than women in an algebra-based physics course for life science majors. As such, while some variation exists, men

express higher self-efficacy toward the generally male-dominated domains of mathematics, physics, and engineering. The differences in self-efficacy are less consistent in more gender equitable biology and chemistry courses.

Differences in the way men and women interpret class achievement information may partially explain these differences in self-efficacy (Marshman et al. 2018). Men interpret their grades and performance in STEM differently than do women (Vermeer et al. 2000), thus providing additional evidence of a differing relationship of the sources of self-efficacy (mastery experiences) by gender. Men may perceive receiving a score of 70% on a test as evidence that they are passing and have strong abilities in science, while women earning a score of 80% on the same test may perceive this as evidence that they lack science abilities (Zimmerman 2000). Thus, a "confidence gap" exists between men and women *despite* comparable, and in some cases, higher prior achievement for women (Kost et al. 2009; Larose et al. 2006). This is particularly important because it is hypothesized that the primary sources of STEM self-efficacy for women include vicarious experience and social persuasion, or others' judgments, feedback, and support, whereas for men, personal mastery experiences are most relevant (Sawtelle et al. 2012a, b; Zeldin et al. 2008; Zeldin and Pajares 2000). The strength of the effect of each source on self-efficacy also appears to vary by age and domain. A recent meta-analysis by Byars-Winston et al. (2017) concluded that for college students, STEM self-efficacy beliefs are more stable and less influenced the four sources of self-efficacy than are the beliefs of K-12 students.

Longitudinal Studies of Self-Efficacy While SCT is intrinsically a dynamic recursive theory where past performance influences current self-efficacy beliefs, relatively little research has measured self-efficacy at multiple time points to investigate the evolution of self-efficacy. Gender differences in STEM self-efficacy at all levels of education have been extensively studied, yet the literature on the *development* of differences in STEM self-efficacy is sparse. In one such 12-year longitudinal study of the development of self-efficacy from grades 1 to 12, Jacobs et al. (2002) found that there was an initial gap in mathematics self-efficacy at grade 1 with boys reporting higher self-efficacy than girls; this gap narrowed linearly until no gap was observed in grade 12. Dramatically different evolution of self-efficacy toward language arts was observed with no gap in grade 1, but with non-linear changes producing a broadening gap which was maximum in grade 9 and closed somewhat by grade 12. The language arts self-efficacy of girls reached a relatively constant value in grade 9. Larose et al. (2006) examined longitudinal changes in self-efficacy as students matriculated from high school to science programs in college and found that 70% of the students did not experience a change in self-efficacy; however, domain- and gender-dependent differences in the patterns of change of self-

efficacy were detected. Caprara et al. (2011) measured the influence of self-efficacy and personality traits on academic performance with a multi-point measurement encompassing junior and senior high school and found self-efficacy beliefs exerted a stronger influence on academic performance later in high school. Pajares and Graham (1999) examined changes in self-efficacy across the academic year for middle-school students and found a decrease in self-efficacy that was attributable to differences in examination difficulty. Very little research has examined the evolution of self-efficacy at the college level.

Theoretical Framework For this work, we employ Bandura's SCT and seek to understand interactions implied by the recursive and multi-threaded (domain specific) features of the theory. By the time students enter college science and mathematics classes, they have had extensive academic experience and the recursive cycle of self-efficacy adjustment has been functioning for many years. As such, students have general expectations of their ability to perform within STEM classes, referred to as STEM self-efficacy in the current study. When students encounter a new STEM academic experience, such as their first college physics or calculus class, they form self-efficacy beliefs toward the specific class, referred to as current class self-efficacy, which according to SCT should be informed by one's STEM self-efficacy. Furthermore, students' current class self-efficacy should evolve in response to performance feedback received in the class; the primary feedback in the classes studied is test scores. Finally, SCT posits that a similar recursive model should operate within each of the classes a student is concurrently taking, creating a multi-stranded model with multiple current class self-efficacies interacting with each other and an evolving STEM self-efficacy. While SCT explicitly posits an evolving self-efficacy, it provides little guidance about the form of the evolution. Presumably, the evolution is non-linear with greater changes the first time one accomplishes a task and lesser changes the 100th time the task is accomplished. College mathematics and science students have been processing STEM mastery experiences into their self-efficacy for over a decade; thus, it is possible that STEM self-efficacy is no longer rapidly changing.

Research Questions This work will present the results of two studies. The methodology of each study is summarized in Fig. 1.

- RQ1 (study 1): Does student self-efficacy vary by STEM academic domain? If so, does the variation depend on gender or subject?
- RQ2 (study 1): Does current class self-efficacy and STEM self-efficacy change over the course of two semesters of physics classes? If so, is this change moderated by gender or academic performance?

- RQ3 (study 2): Is the gender difference in current class self-efficacy observed in Study 1 present very early in the class?
- RQ4 (study 2): Does current class self-efficacy change as a result of class feedback? If so, is this change moderated by gender?

For RQ1, we hypothesized based on the generally lower levels of STEM self-efficacy reported for women in the literature, that women would report lower levels of self-efficacy across all STEM domains. Because of research demonstrating different modes of development of self-efficacy by gender, we hypothesized for RQ2 that the change in self-efficacy between classes would be different by gender, with the change of self-efficacy for men more influenced by class feedback (mastery experiences). RQ3 and RQ4 represent an exploratory analysis which was initiated when the results of RQ1 did not support our hypotheses; therefore, no directional hypothesis was made for these questions.

Methods

Study 1 Sample This study was conducted from the fall 2015 to the spring 2017 semesters in introductory calculus-based physics courses and introductory calculus courses at a large eastern US land-grant university serving approximately 30,000 students. Students from the introductory calculus-based mechanics course, Physics 1 (Phys 1), and introductory calculus-based electricity and magnetism course, Physics 2 (Phys 2), were included in the study, as were students from two Calculus 1 course sequences for STEM majors: Calculus 1 (Cal 1) and Stretch Calculus 1 (SCal 1), which combines pre-calculus and calculus into a two-semester sequence taken by students not academically prepared to enter the one semester Cal 1.

Data were collected in both Phys 1 and Phys 2 for the 2-year period from the fall 2015 semester to the spring 2017 semester. This data set was analyzed both in aggregate and by course. For the period studied, 3266 students enrolled in the two courses of which 3083 completed the courses. Students repeating the course or who did not take one of the in-semester examinations were removed. Of the remaining students, 1896 completed the self-efficacy survey; these students form the physics data set for study 1. Overall, the students were 77% men and 23% women; 89% were pursuing engineering majors. Race/ethnicity data provided by the institution showed the students were 86% White, 3% African American, 5% Asian, and 2% Hispanic; 3% of the students reported multiple races or ethnicities. Disaggregating by course showed Phys 1 with $N = 1012$ students and Phys 2 with $N = 884$ students. Most engineering majors are required to

Fig. 1 Research plan. SEF self-efficacy

Study 1	Study 2
<u>Research Question 1</u>	<u>Research Question 3</u>
<p>Classes: SCal 1, Cal 1, Phys 1, Phys 2</p> <p>Measures: Current Class SEF, Math Classes SEF, Science Classes SEF, Major Classes SEF, Professional SEF</p> <p>Duration: Fall 2015 to Spring 2017</p> <p>Application: (Cross-Sectional) One Survey Mid-Semester</p>	<p>Classes: Phys 1, Phys 2</p> <p>Measures: Current Class SEF</p> <p>Duration: Fall 2016 to Spring 2017</p> <p>Application: One Survey Early in the Semester</p>
<u>Research Question 2</u>	<u>Research Question 4</u>
<p>Classes: Phys 1, Phys 2</p> <p>Measures: Current Class SEF, Math Classes SEF, Science Classes SEF, Major Classes SEF, Professional SEF, STEM SEF, Test Ave, GPA</p> <p>Duration: Fall 2015 to Spring 2017</p> <p>Application: (Repeated Measures) One Survey Mid-Semester in Each Class</p>	<p>Classes: Phys 1, Phys 2</p> <p>Measures: Current Class SEF, Test Ave</p> <p>Duration: Fall 2016 to Spring 2017</p> <p>Application: Two Surveys: (Repeated Measures) One Early in the Semester, One Mid-Semester</p>

take both Phys 1 and 2, and as such, the expected enrollment pattern is for students to take Phys 2 the semester after Phys 1. Many students matriculated through both courses. The two courses were analyzed in aggregate and were also disaggregated by course. To form the aggregated data set, all self-efficacy measures and test averages were averaged over the two courses leaving $N = 1557$ distinct participants. There were a subset of students for which data was available for two semesters allowing the investigation of the evolution of self-efficacy with time. A total of $N = 339$ students enrolled in Phys 1 later enrolled in Phys 2 and completed the surveys in both classes.

Data were also collected from calculus courses in the fall 2015, fall 2016, and spring 2017 semesters. For Cal 1, 1694 students enrolled in the course, 1499 students completed the course with a grade, and 904 students completed the self-efficacy survey [men (72%); engineers (74%); white (87%)]. For SCal 1, 1144 students enrolled in the class, 1079 students completed the class with a grade, and 528 students completed the self-efficacy survey [men (57%); engineers (49%); white (84%)]. No students were enrolled in both calculus classes.

Study 2 Sample When the hypothesized differences in self-efficacy by gender were not observed in preliminary analyses, an additional self-efficacy instrument was administered early in the semester in the physics classes in the fall 2016 and spring 2017 semesters. Students in the sample for study 1 who also completed these additional survey items form the data set for study 2. Of the 911 students for sample 1 enrolled in the fall 2016 or spring 2017 semesters, $N = 815$ completed the second current class self-efficacy survey and form the data set for study 2 [men (75%); engineers (89%); white (86%)]. As with study 1, this data set will be

analyzed in aggregate and by course [Phys 1, $N = 453$; Phys 2, $N = 362$]. When analyzed in aggregate, multiple instances of the same student will be averaged producing a data set with $N = 768$ unique students.

Measures Because self-efficacy is domain specific (Bandura 1997; Pajares and Miller 1995), a survey was designed to assess self-efficacy beliefs in the current physics or mathematics class (current class self-efficacy), other science classes (science classes self-efficacy), other mathematics classes (mathematics classes self-efficacy), classes within the student's major department (major classes self-efficacy), and functioning within the student's planned future profession (professional self-efficacy). Each of these subscales was developed using a modified version of the "Self-Efficacy for Learning and Performance" subscale of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al. 1991). This instrument has been extensively validated (Pintrich et al. 1993) and widely used (Duncan and McKeachie 2005). In a recent study exploring the latent factor structure of the MSLQ, the self-efficacy subscale was one of two subscales of the MSLQ demonstrating a Cronbach's alpha of .90 or higher (Hilpert et al. 2013). The MSLQ measures a general sense of student self-efficacy within a class. To generate questions measuring self-efficacy in different domains, the domain name was substituted for "class/course" into a six-question subset of the original eight question MSLQ self-efficacy subscale. The full original scale and the modified scale for the physics classes is presented in Table 1. The original question numbers are prefixed with MSL, the modified questions with A. For example, modified item 5 is A_Q5. Questions are numbered consistently with Pintrich et al. (1993).

Table 1 Physics self-efficacy scale. The original MSLQ self-efficacy subscale and its modifications for the physics domain

MSLQ #	MSLQ Item	Adapted #	Revised Item
MSL_Q5	I believe that I will receive an excellent grade in this class.	A_Q5	I believe that I will receive an excellent grade in this physics class
MSL_Q6	I'm certain I can understand the most difficult material presented in the readings for this course.	A_Q6	Not used.
MSL_Q12	I'm confident I can understand the basic concepts taught in this course.	A_Q12	I'm confident I can understand the basic concepts taught in this physics course.
MSL_Q15	I'm confident I can understand the most complex material presented by the instructor in this course.	A_Q15	I'm confident I can understand the most complex material presented by the instructor in this physics course.
MSL_Q20	I'm confident I can do an excellent job on the assignments and tests in this course.	A_Q20	I'm confident I can do an excellent job on the assignments and tests in this physics course.
MSL_Q21	I expect to do well in this class.	A_Q21	Not used.
MSL_Q29	I'm certain I can master the skills being taught in this class.	A_Q29	I'm certain I can master the skills being taught in this physics class.
MSL_Q31	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.	A_Q31	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this physics class.

Responses used a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.”

The method of word substitution into the MSLQ was also employed by Vogt et al. (2007) to investigate self-efficacy toward engineering majors. The scale was compressed from 8 to 6 questions to shorten the final instrument as it was administered as part of a larger package of surveys; the high Cronbach alpha value of .93 and .94 reported in Pintrich et al. (1993) and Hilpert et al. (2013), respectively, suggested that a shorter scale could still provide high reliability. The original MSL_Q6, which asked about reading assignments, was discarded because college STEM classes require substantially more from students than reading comprehension. MSL_Q21 is a shorter version of MSL_Q31 and was also removed. Finally, the self-efficacy toward their intended profession subscale did not include MSL_Q12 because mastering only the basic concepts toward one’s profession seemed unnatural. Some slight rephrasing was performed when necessary; for example, for the professional domain in MSL_Q5 “excellent grade” was changed to “excellent performance reviews.” The 35 questions were randomly ordered and combined with 8 additional questions asking about career decision making as part of a larger effort to understand retention.

The newly adapted self-efficacy questions were closely related to the original questions in the MSLQ and were therefore closely related to questions in a highly validated and tested instrument. The new instrument went through a three-stage validation process during the summer 2015 semester. The process began with 16 structured 30-min interviews with students in the summer Phys 1 and 2 classes. Survey validation continued by administering open-response versions of the survey questions to 36 students. Students were asked to restate the questions and to suggest alternate phrasing. Student responses

to the structured interviews and the open-response surveys were then used to modify the self-efficacy assessment for each domain. Data were collected in fall 2015 and spring 2016 with the modified instruments to complete the validation process. Internal scale reliability was characterized by calculating Cronbach’s alpha for each domain subscale and showed that each subscale was reliable at above the minimum .70: physics classes (.93), mathematics classes (.93), other science classes (.92), classes in the major (.89), and the profession (.87).

In addition to the self-efficacy survey, demographic information including gender, race/ethnicity, and college grade point average (GPA) was collected from university records. The student’s test scores in the physics classes were provided by the instructor and are reported as percentages (TestAve).

Instructional Environment The physics classes were presented with three 50-min lectures and one required 3-h laboratory each week. Each course was taught by two instructors in multiple lecture sections each semester. The instructors made substantial efforts to include research-based pedagogical elements in the learning experience including peer instruction with clickers in the lecture and inquiry-based learning in the laboratory.

There were two variants of Calculus 1: a traditional one-semester course (Cal 1) and an equivalent stretch two-semester course (SCal 1) for students not academically prepared to take Cal 1. Cal 1 met for three 50-min lectures and two 75-min recitations sessions each week, while SCal 1 met for three 50-min lectures and one 50-min recitation session each week. Both classes were taught by multiple instructors who employed a variety of research-based pedagogies including group learning and problem-based instruction to develop strong conceptual understanding.

Analysis Methods All research questions were investigated with linear mixed effects (lme) modeling. Students were treated as random effects to control for within-subject correlation. Linear mixed-effect models compensate for the imbalanced sample size caused by the underrepresentation of women in physics and mathematics classes. To account for clustering within courses, all analyses were repeated disaggregating the students by course. All statistical analysis was performed with the “R” software package. The “lme4” package (Bates et al. 2014) was employed for linear mixed effects modeling.

Study 1 Results

Study 1 measured the self-efficacy of mathematics and physics students toward multiple STEM domains for the 2-year period from fall 2015 to spring 2017. Descriptive statistics for the four courses are presented in Table 2 and disaggregated by gender in Table 3. The correlation matrix for the physics courses aggregating the two courses (Phys 1 and Phys 2) is presented in Table 4. Mean scores were high across all domains, but professional self-efficacy was consistently higher than the other domains. As expected, the correlations between the various types of self-efficacy were also high.

RQ1: Does student self-efficacy vary by STEM academic domain? If so, does the variation depend on gender or subject? Differences of self-efficacy in the physics classes by domain were investigated using linear mixed effects modeling (lme). This analysis was performed aggregating Phys 1 and Phys 2. Self-efficacy by domain was converted to a percent of maximum possible (POMP) scale, thus projecting the [1, 5] range of the Likert scale onto the [0, 100] range to allow its interpretation as a percent change. The lme model is shown in Eqs. 1 and 2. The overall self-efficacy of student *i* in domain *j* (SelfEfficacy_{*ij*}) is the dependent variable. The independent variables are dummy-coded variables representing the domain: math class self-efficacy (MC), science class self-efficacy (SC), major class self-efficacy (MJ), and professional self-efficacy (PR). For example, MC_{*j*} is one if the domain is other math classes

and zero otherwise. The physics class self-efficacy was used as the reference level.

$$\text{SelfEfficacy}_{ij} = \beta_{0i} + \beta_1 \times MC_j + \beta_2 \times SC_j + \beta_3 \times MJ_j + \beta_4 \times PR_j + \varepsilon_{ij} \tag{1}$$

$$\beta_{0i} = \gamma_0 + u_i \tag{2}$$

Correlations resulting from multiple measures of the same student were modeled with a random intercept, β_{0i} . Equation 2 models this intercept with an overall mean, γ_0 , and a random effect for each student, u_i . The residual errors are assumed to be normally distributed, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma)$, as are the student random effects, $u_i \sim \mathcal{N}(0, \tau)$. A lme model including domain significantly improved goodness of fit over the model including only the intercept using a likelihood ratio test [$\chi^2(4) = 1013, p < .001$]. The full set of regression coefficients with confidence intervals as well as the variance of the error and the random effects is reported in Table 5. Equation 2 will be used to model the random intercepts in all models in this work.

Post hoc analysis using paired *t* tests with a Bonferroni correction showed current class self-efficacy was significantly different than all other domains ($ps < .001$) with effect sizes from $d = .32$ to $d = .57$. Professional self-efficacy was also significantly different from all other domains ($ps < .001$) with effect sizes from $d = .20$ to $d = .57$. There was no statistically significant difference between science class self-efficacy, mathematics class self-efficacy, and major class self-efficacy. As such, students do not strongly differentiate their self-efficacy expectations toward other STEM classes. Because of the similarity between these variables and their high correlation ($r = .79$ to $.86$), science class self-efficacy, mathematics class self-efficacy, and major class self-efficacy were averaged to form a general STEM self-efficacy (STEM self-efficacy).

This analysis was repeated in the aggregated mathematics courses with similar results. The model including domain (Eq. 1) significantly improved goodness of fit over the model including only the intercept using a likelihood ratio test [$\chi^2(4) = 493.37, p < .001$]. The results of this analysis are reported in Table 6.

Table 2 Domain-specific self-efficacy. Self-efficacy was measured on a 5-point Likert scale. Calculus 1 is a traditional single semester calculus course. S. Calculus 1 is the two-semester stretch Calculus 1

Course	Number	Current class self-efficacy (M ± SD)	Science class self-efficacy (M ± SD)	Mathematics class self-efficacy (M ± SD)	Major class self-efficacy (M ± SD)	Professional self-efficacy (M ± SD)
S. Calculus 1	528	3.76 ± .80	3.90 ± .70	3.88 ± .68	3.94 ± .67	4.08 ± .64
Calculus 1	904	3.66 ± .79	3.90 ± .71	3.92 ± .64	3.91 ± .64	3.99 ± .62
Physics 1	1012	3.81 ± .85	4.05 ± .69	4.09 ± .70	4.09 ± .67	4.25 ± .62
Physics 2	884	3.91 ± .79	4.12 ± .67	4.19 ± .67	4.17 ± .63	4.26 ± .61

Table 3 Domain-specific self-efficacy for men and women. Self-efficacy was measured on a 5-point Likert scale. Calculus 1 is a traditional single semester calculus course. S. Calculus 1 is the two-semester stretch Cal 1

Course	Number	Current class self-efficacy (M ± SD)	Science class self-efficacy (M ± SD)	Mathematics class self-efficacy (M ± SD)	Major class self-efficacy (M ± SD)	Professional self-efficacy (M ± SD)
Men						
S. Calculus 1	304	3.88 ± .75	3.93 ± .68	3.94 ± .66	3.94 ± .67	4.06 ± .66
Calculus 1	657	3.71 ± .76	3.90 ± .71	3.93 ± .65	3.93 ± .65	3.98 ± .63
Physics 1	749	3.90 ± .80	4.07 ± .68	4.09 ± .70	4.12 ± .66	4.25 ± .62
Physics 2	706	3.94 ± .78	4.12 ± .66	4.20 ± .66	4.19 ± .62	4.27 ± .61
Women						
S. Calculus 1	224	3.60 ± .84	3.86 ± .74	3.80 ± .71	3.95 ± .68	4.09 ± .63
Calculus 1	247	3.55 ± .85	3.90 ± .70	3.89 ± .62	3.85 ± .63	4.00 ± .57
Physics 1	263	3.56 ± .92	4.01 ± .73	4.07 ± .71	4.03 ± .70	4.23 ± .63
Physics 2	178	3.78 ± .81	4.09 ± .69	4.18 ± .70	4.11 ± .66	4.25 ± .63

Post hoc analysis with a Bonferroni correction showed that current class self-efficacy was significantly different from all other domains ($ps < .001$) with effect sizes from $d = .27$ to $d = .45$. Professional self-efficacy was also significantly different from all other domains [current class self-efficacy, $p < .001$, $d = .45$, science class self-efficacy, $p < .001$, $d = .18$; mathematics class self-efficacy, $p < .001$, $d = .18$; major class self-efficacy, $p = .001$, $d = .15$]. There were no significant differences between science class self-efficacy, mathematics class self-efficacy, and major class self-efficacy.

Within the physics courses, the role of domain and gender was investigated for current class self-efficacy, STEM self-efficacy, and professional self-efficacy using lme modeling to fit the model shown in Eq. 3 where Gen_i is zero for women and one for men, PR_j is one if the domain is professional self-efficacy, and CC_j is one if the domain is current class self-efficacy. STEM self-efficacy was used as the reference level. Equation 3 presents the final model after non-significant interactions were removed. The full regression results are reported in Table 7.

$$\begin{aligned}
 \text{SelfEfficacy}_{ij} = & \beta_{0i} + \beta_1 \times CC_j + \beta_2 \times PR_j + \beta_3 \\
 & \times Gen_i + \beta_4 \times CC_j \times Gen_i + \beta_5 \times PR_j \\
 & \times Gen_i + \varepsilon_{ij} \tag{3}
 \end{aligned}$$

Table 4 Correlation matrix for aggregated physics classes. All $ps < .001$. Current class refers to the current physics class

	1	2	3	4	5
1. Current class self-efficacy	–				
2. Science class self-efficacy	.70	–			
3. Math class self-efficacy	.59	.82	–		
4. Major class self-efficacy	.65	.86	.81	–	
5. Professional self-efficacy	.56	.79	.74	.82	–

This analysis was performed aggregating Phys 1 and Phys 2. Current class self-efficacy was converted to a POMP scale thus projecting the [1, 5] range of the Likert scale onto the [0, 100] range to allow its interpretation as a percent change. A lme model including domain but not gender significantly improved goodness of fit over the model including only the intercept using a likelihood ratio test [$\chi^2(2) = 768$, $p < .001$]. Professional self-efficacy was significantly higher than STEM self-efficacy [$B = 3.41$, $SE = .36$; $t(3114) = 9.56$, $p < .001$], and current class self-efficacy was significantly lower than STEM self-efficacy [$B = -6.92$, $SE = .36$; $t(3114) = 19.4$, $p < .001$]. The full model that included both domain and gender significantly improved goodness of fit over the model without gender [$\chi^2(3) = 88.3$, $p < .001$]. Both current class self-efficacy [$B = -11.4$, $SE = .73$; $t(3114) = 15.6$, $p < .001$] and professional self-efficacy [$B = 4.07$, $SE = .73$; $t(3114) =$

Table 5 Linear mixed effect model for Eq. 1 in the aggregated physics classes. Dependent variable self-efficacy with dummy-coded domain-independent variables. Current physics class self-efficacy was used as the reference level

Predictors	Estimates	CI	p
Intercept	70.72	69.87–71.57	< .001
Math class self-efficacy	7.56	6.93–8.20	< .001
Science class self-efficacy	6.01	5.37–6.65	< .001
Major class self-efficacy	7.20	6.56–7.83	< .001
Professional self-efficacy	10.33	9.69–10.97	< .001
Random effects			
σ^2	82.64		
τ^2	211.87		
ICC	.72		
N_{subj}	1557		

Table 6 Linear mixed effect model for Eq. 1 in the aggregated math classes. Dependent variable self-efficacy with dummy-coded domain independent variables. Current math class self-efficacy was used as the reference level

Predictors	Estimates	CI	<i>p</i>
Intercept	67.46	66.55–68.38	< .001
Math class self-efficacy	5.16	4.45–5.87	< .001
Science class self-efficacy	5.04	4.33–5.76	< .001
Major class self-efficacy	5.55	4.84–6.27	< .001
Professional self-efficacy	8.03	7.32–8.75	< .001
Random effects			
σ^2	95.17		
τ^2	205.58		
ICC	.68		
N_{subj}	1365		

5.56, $p < .001$] were significantly different than STEM self-efficacy. The gender-by-current-class-self-efficacy interaction was also significant with men having higher current class self-efficacy than women [$B = 5.85$, $SE = .84$; $t(3114) = 7.01$, $p < .001$]. The main effect of gender and the gender-by-professional-self-efficacy interactions were not significant. Post hoc analysis with a Bonferroni correction showed there was no significant difference between the self-efficacy of men and women except for current class self-efficacy [$t(547) = 5.85$, $p < .001$, $d = .37$]. Similar results were obtained when the models were fit separately for the individual courses.

An analysis of both domain and gender in the aggregated mathematics courses yielded similar results. The model including domain and gender was a significant improvement over the model including only domain [$\chi^2(3) = 61.9$, $p < .001$]. Current class self-efficacy was significantly lower than STEM self-efficacy [$B = -7.58$, $SE = .61$; $t(2928) = 12.4$, $p < .001$], and professional self-efficacy was significantly higher than STEM self-efficacy [$B = 4.25$, $SE = .61$;

$t(2928) = 6.96$, $p < .001$]. The main effect of gender was not significant. However, the gender-by-current-class-self-efficacy interaction was significant with men expressing higher current class self-efficacy than women [$B = 3.48$, $SE = .75$; $t(2928) = 4.66$, $p < .001$], but unlike in the physics classes, the gender-by-professional-self-efficacy interaction was significant with women expressing higher professional self-efficacy than men [$B = -2.20$, $SE = .75$; $t(2928) = 2.95$, $p = .003$]. The full regression results are shown in Table 8.

Post hoc analysis with a Bonferroni correction showed there was no significant difference between the self-efficacy of men and women except in current class self-efficacy [$t(852) = 4.08$, $p < .001$, $d = .24$]. Similar results were obtained disaggregating by course except that the gender-by-professional-self-efficacy interaction was not significant in Cal 1.

Summary

The progression of self-efficacy was the same in all courses with lowest self-efficacy in the current class, similar and higher self-efficacy toward other mathematics and science courses and courses in the major, and highest self-efficacy toward their intended profession. For students in all classes measured, general STEM self-efficacy was the same for men and women, while women expressed lower self-efficacy toward the mathematics or science class they were currently taking than men.

RQ2: Does current class self-efficacy and STEM self-efficacy change over the course of two semesters of physics classes? If so, is this change moderated by gender or academic performance? Table 9 presents the mean self-efficacy for students who matriculated through both physics courses. Self-efficacy changed little between the two courses. Pairwise t tests with a Bonferroni correction showed no significant differences between courses for any domain for men. For women, the increase in current class self-efficacy [$t(80) = 2.91$,

Table 7 Linear mixed effect model for Eq. 3 in the aggregated physics classes. Dependent variable self-efficacy with dummy-coded domain-independent variables and gender. STEM self-efficacy was used as the reference level

Predictors	Estimates	CI	<i>p</i>
Intercept	76.42	74.65–78.19	< .001
Gender	1.60	-.42–3.61	.121
Current class self-efficacy	-11.42	-12.86– -9.99	< .001
Professional self-efficacy	4.07	2.64–5.51	< .001
Current class self-efficacy \times gender	5.85	4.22–7.49	< .001
Professional self-efficacy \times gender	-0.86	-2.50– .78	.302
Random effects			
σ^2	96.65		
τ^2	196.74		
ICC	.67		
N_{subj}	1557		

Table 8 Linear mixed effect model for Eq. 3 in the aggregated math classes. Dependent variable self-efficacy with dummy-coded domain-independent variables and gender. STEM self-efficacy was used as the reference level

Predictors	Estimates	CI	<i>p</i>
Intercept	71.88	70.31–73.44	< .001
Gender	1.23	-.68–3.15	.206
Current class self-efficacy	-7.58	-8.78–-6.39	< .001
Professional self-efficacy	4.25	3.05–5.45	< .001
Current class self-efficacy × gender	3.48	2.01–4.94	< .001
Professional self-efficacy × gender	-2.20	-3.66–-.73	.003
Random effects			
σ^2	87.84		
τ^2	202.97		
ICC	.70		
N_{subj}	1365		

$p = .023, d = .33$] was significant where a Bonferroni correction has been applied.

This question was further investigated with a lme model using the two time-separated measurements of self-efficacy as a within-subject measurement. The full model is shown in Eq. 4. The current class self-efficacy ($CCSE_{ij}$) of student i at time j was measured on a POMP scale. Current class self-efficacy was not normalized because the means at two time points were being compared.

$$CCSE_{ij} = \beta_{0i} + \beta_1 \times \text{Time}_j + \beta_2 \times \text{Gen}_i + \beta_3 \times \text{TestAve}_{ij} + \varepsilon_{ij} \tag{4}$$

The model including time (Phys 1, Phys 2) was a significant improvement over the model including only the intercept indicating that current class self-efficacy did change with time [$\chi^2(1) = 4.64, p = .031$]. Time was coded with Phys 1 as 0, Phys 2 as 1 and had regression coefficient [$B = 2.26, SE = 1.05; t(339) = 2.16, p = .031$]; a student’s average self-efficacy increased 2% when matriculating from Phys 1 to Phys 2. A

model including time and gender was a significant improvement over the model including time alone [$\chi^2(2) = 9.72, p = .008$]. In this model, there was a significant main effect of gender [$B = 7.19, SE = 2.40; t(548) = 2.99, p = .003$], a significant main effect of time [$B = 6.53, SE = 2.12; t(339) = 3.08, p = .002$], and a significant time-by-gender interaction with women with higher current class self-efficacy in Phys 2 [$B = -5.61, SE = 2.43; t(339) = 2.31, p = .022$]. To explore the effect of student performance on students’ longitudinal self-efficacy, test average was added to the model. For this analysis, test averages were normalized separately for each course to eliminate the overall effect of a small difference in test average between the courses. A saturated model including time, gender, and TestAve and all interactions significantly improved on the model without TestAve [$\chi^2(4) = 209, p < .001$]. Removing interactions that were not significant did not significantly change model fit. The final model demonstrated significant main effects of gender [$B = 5.66, SE = 1.74; t(339) = 3.26, p = .001$], time [$B = 2.26, SE = .91; t(339) = 2.47, p = .014$], and TestAve [$B = 10.3, SE = .65; t(634) = 15.8, p < .001$]. There were no significant interactions. The full regression results are shown in Table 10. As such, current class self-efficacy increased 2.26% for students successfully matriculating from Phys1 to Phys 2 correcting for test average; this change was the same for both men and women.

The change in STEM self-efficacy with time was also investigated between Phys 1 and Phys 2. This variable was expected to be related to more general measures of academic performance such as cumulative college GPA and was expected to be more stable than domain-specific self-efficacy; college GPA (CGPA) was not normalized because two time points were being compared. A lme model with STEM self-efficacy (STEMSE) as the dependent variable showed that this was the case (Eq. 5): STEM self-efficacy was a very stable construct through Phys 1 and Phys 2.

$$\text{STEMSE}_{ij} = \beta_{0i} + \beta_1 \times \text{Time}_j + \beta_2 \times \text{Gen}_i + \beta_3 \times \text{CGPA}_{ij} + \varepsilon_{ij} \tag{5}$$

Table 9 Students matriculating through both Physics 1 and Physics 2. Self-efficacy was measured on a 5-point Likert scale

Course	Number	Current class self-efficacy ($M \pm SD$)	Science class self-efficacy ($M \pm SD$)	Mathematics class self-efficacy ($M \pm SD$)	Major class self-efficacy ($M \pm SD$)	Professional self-efficacy ($M \pm SD$)
Men						
Physics 1	258	4.00 ± .75	4.14 ± .65	4.11 ± .67	4.19 ± .60	4.32 ± .56
Physics 2		4.04 ± .73	4.14 ± .65	4.20 ± .63	4.22 ± .58	4.28 ± .57
Women						
Physics 1	81	3.71 ± .91	4.13 ± .71	4.12 ± .71	4.14 ± .64	4.31 ± .59
Physics 2		3.97 ± .68	4.19 ± .64	4.22 ± .68	4.11 ± .67	4.34 ± .56

Table 10 Linear mixed effect model for Eq. 4 examining longitudinal changes between physics classes. The dependent variable is current class self-efficacy with independent variables gender (women = 0, men = 0), time (Phys 1 = 0, Phys 2 = 1), and test average

Predictors	Estimates	CI	<i>p</i>
Intercept	68.96	65.86–72.06	< .001
Time	2.26	.47–4.05	.013
Gender	5.66	2.26–9.07	.001
Test average	10.29	9.01–11.57	< .001
Random effects			
σ^2	141.65		
τ^2	114.62		
ICC	.45		
N_{subj}	339		

There was no significant effect of class, college GPA, nor was there a significant interaction. As such, while a student's current class self-efficacy changes as a result of class performance feedback (RQ4) and increases between physics classes, STEM self-efficacy appears not to evolve with time even when correcting for overall academic performance measured by college GPA.

Summary

Overall, the student's self-efficacy toward his or her current physics class increased somewhat as he or she matriculated from Phys 1 to Phys 2; however, self-efficacy toward STEM classes in general did not change over the same time period. Neither the change in current class self-efficacy or STEM self-efficacy was moderated by gender or academic performance.

Study 2 Results

The pattern of gender differences in self-efficacy measured in study 1 was not expected. The hypothesized lower self-efficacy of women toward STEM domains was only observed in their current mathematics or physics class; men and women reported equal self-efficacy toward other STEM domains. While research into STEM self-efficacy has shown mixed results for the differences in the self-efficacy of men and women, we expected the differences to be fairly consistent across the closely related STEM domains of mathematics, physics, and engineering (major classes). To further explore the unexpected differences between current class self-efficacy and self-efficacy in other domains, a second survey measuring current class self-efficacy was administered early in the semester in the fall 2016 and spring 2017 physics courses before substantive feedback had been provided to the students. This provided

a measurement of current class self-efficacy at two time points: t_1 early in the semester and t_2 in the middle of the semester after the second test.

RQ3: Is the gender difference in current class self-efficacy observed in study 1 present very early in the class? Early in the semester, a significant gap in current class self-efficacy (t_1) was already present between men and women [men ($N = 581$), $M = 4.10$, $SD = .65$; women ($N = 187$), $M = 3.74$, $SD = .79$; $t(271) = 5.61$, $p < .001$, $d = .52$]. The .36 gap early in the semester closed somewhat by the second measurement of current class self-efficacy (t_2) to the .27 gap observed mid-semester. The mid-semester difference of current class self-efficacy (t_2) was also significant [men ($N = 581$), $M = 4.03$, $SD = .74$; women ($N = 187$), $M = 3.76$, $SD = .82$; $t(290) = 4.00$, $p < .001$, $d = .35$]. These results are reported using the original 5-point Likert scale. The self-efficacy survey given early in the semester was administered before the students received any substantive feedback in the form of homework, quiz, or test grades; as such, the differences in the current class self-efficacy of men and women observed in study 1 do not result from the effect of feedback from the current course.

RQ4: Does current class self-efficacy change as a result of class feedback? If so, is this change moderated by gender? The time evolution of current class self-efficacy in the physics courses was investigated with lme models. The current class self-efficacy (CCSE_{*ij*}) of student *i* at time *j* was modeled using Eq. 6. This is the final model after non-significant interactions have been removed.

$$\text{CCSE}_{ij} = \beta_{0i} + \beta_1 \times \text{Time}_j + \beta_2 \times \text{Gen}_i + \beta_3 \times \text{TestAve}_i + \beta_4 \times \text{TestAve}_i \times \text{Time}_j + \varepsilon_{ij} \quad (6)$$

Aggregating the two physics courses, a lme model treating students as random effects and including the observations of current class self-efficacy at two time points did not improve model fit over a model with only the intercept. Current class self-efficacy was measured on a POMP scale, and TestAve was normalized. Time was coded with early in the semester as zero (t_1), mid-semester as one (t_2). Test average was only available at t_2 . Current class self-efficacy was not normalized because two time points were being compared. The main effect of time was not significant: on average, current class self-efficacy changed little within the physics classes. A model that included both time and gender was a significant improvement over a model including only time [$\chi^2(2) = 36.3$, $p < .001$] with gender as a significant main effect [$B = 8.97$, $SE = 1.53$; $t(1209) = 5.87$, $p < .001$]. Neither time nor the gender-by-time interaction was significant. A model including TestAve, time, and gender was a significant improvement over the model including only time and gender. The saturated model including all interactions significantly improved model fit over the model with only time and gender [$\chi^2(4) = 238$, $p < .001$].

Removing terms where the regression coefficients were not significant yielded a model that was not significantly different than the saturated model. The final model included a significant main effect of gender [$B = 7.34$, $SE = 1.21$; $t(768) = 6.05$, $p < .001$] and a significant main effect of TestAve [$B = 3.55$, $SE = .60$; $t(1234) = 5.89$, $p < .001$]. The main effect of time was not significant. There was a significant time-by-TestAve interaction [$B = 5.98$, $SE = .61$; $t(768) = 9.88$, $p < .001$] interaction. The time-by-gender, gender-by-TestAve, and the three-way interaction between gender, time, and TestAve were not significant; TestAve had an equal effect on self-efficacy for men and women. The final model is summarized in Fig. 2 and the regression results in Table 11. The main effect of TestAve captures the general relation of class performance and current class self-efficacy. The time-by-TestAve interaction captures the effect of TestAve on the change in current class self-efficacy with time. While significant, the modification of current class self-efficacy due to class feedback, measured by test average, is practically small. A one standard-deviation increase in TestAve only produced a 5.98-point increase in current class self-efficacy on a 100-point scale.

The final model was also fit for the individual courses. The results were fairly similar except for the time-by-gender interaction. In Phys 1, the time-by-gender interaction was not significant while in Phys 2, the time-by-gender interaction was significant [$B = -5.22$, $SE = 1.97$; $t(362) = 2.64$, $p = .008$].

Summary

Current class self-efficacy did change during a physics class, and this change was moderated by the student’s test average. The change was moderated by gender in Phys 2 but not Phys

Table 11 Linear mixed effect model for Eq. 6 examining changes of self-efficacy within physics classes. The dependent variable is current class self-efficacy with independent variables gender (women = 0, men = 0), time (early in class = 0, mid-semester = 1), and test average

Predictors	Estimates	CI	<i>p</i>
Intercept	69.72	67.57–71.87	< .001
Time	–1.05	–2.24–.14	.083
Gender	7.34	4.97–9.72	< .001
Test average	3.55	2.37–4.73	< .001
Time × test average	5.98	4.80–7.17	< .001
Random effects			
σ^2	140.76		
τ^2	137.56		
ICC	.49		
N_{subj}	768		

1. This resulted in a small narrowing of the gender differences in current class self-efficacy by mid-semester in Phys 2. The gender-by-time-by-TestAve interaction was not significant suggesting that men and women process performance feedback in the form of test scores into their current class self-efficacy equally.

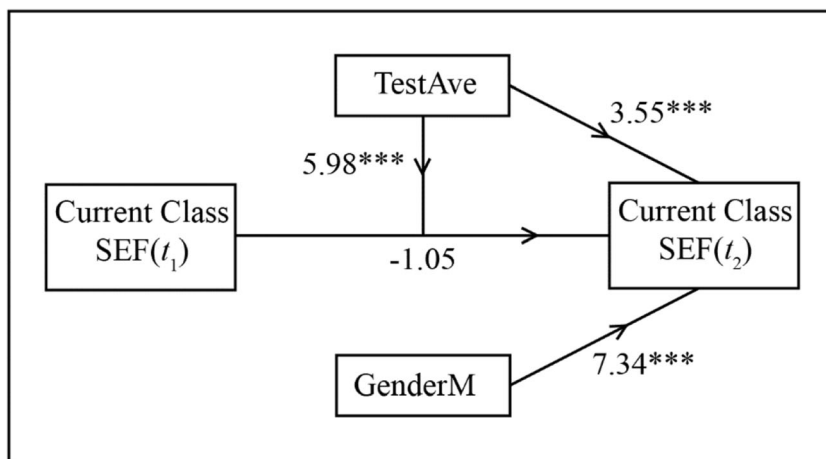
Discussion

The current study explored self-efficacy in multiple STEM domains concurrently. Students’ self-efficacy beliefs were different by academic STEM domain supporting the previous literature that showed that self-efficacy beliefs are multidimensional (Zimmerman 2000). In both physics and mathematics courses, student self-efficacy was the lowest toward the current class and the highest toward the students’ intended profession while self-efficacy toward other science classes, other mathematics classes, and classes in their major were similar, demonstrating that STEM self-efficacy differed by some, but not all, domains. General STEM self-efficacy (the average of science class self-efficacy, mathematics class self-efficacy, and majors class self-efficacy) was also different from self-efficacy toward the current class and the intended profession. These results suggest that students have at minimum three tiers of STEM self-efficacy: self-efficacy toward their current STEM class, a general STEM self-efficacy, and self-efficacy toward their intended STEM profession.

Previous research has shown that women tend to report lower levels of self-efficacy in specific academic STEM domains (Huang 2013; Kost et al. 2009; Marshman et al. 2018). This was only partially supported by this study. There was a significant difference in self-efficacy toward the current mathematics or physics class in which the student was enrolled; however, men and women expressed similar STEM self-efficacy and professional self-efficacy. This observation serves to explain the inconsistencies reported in differences in self-efficacy between men and women; different results would be obtained if the students were asked about STEM classes in general or a class in which they were currently enrolled. This observation also supports the need to ask questions about specific domains or tasks when investigating self-efficacy. This work partially supports to findings of Nissen (2019) who reported a difference in self-efficacy between boys and girls in high school physics, but no difference in mathematics. In college calculus classes, we also find differences in self-efficacy.

Self-Efficacy and Time STEM self-efficacy did not change significantly between Phys 1 and 2, nor did it change in response to academic feedback in the form of cumulative college GPA. Because STEM self-efficacy did not evolve with either time or academic performance, it can be viewed as a

Fig. 2 Optimal model for the evolution of current class self-efficacy within a class. * $p < .05$; ** $p < .01$; *** $p < .001$. SEF self-efficacy



somewhat stable variable that can affect current class self-efficacy but which is little affected by changes in current class self-efficacy. STEM self-efficacy appears to be a stable construct in the early college years after the students have matriculated through Cal 1 (Phys 1 has Cal 1 as a prerequisite). This was consistent with research showing college students have more stable self-efficacy than high school students (Byars-Winston et al. 2017). The stability of STEM self-efficacy with time was consistent with previous research showing that for 70% of students, self-efficacy does not change in the transition from high school to college (Larose et al. 2006). It was also consistent with the saturation of language arts self-efficacy in high school observed by Jacobs et al. (2002), but inconsistent with the continued increase in mathematics self-efficacy in the same study.

Current class self-efficacy increased by 2.26% from Phys 1 to Phys 2 corrected for test average. This result suggests that students' domain-specific self-efficacy continues to evolve as they process academic feedback into their self-efficacy toward the next class in the same subject even though their STEM self-efficacy did not change.

Self-Efficacy and Feedback Many sources of self-efficacy beliefs can influence changes in a student's self-efficacy (Bandura 1977, 1986, 1993, 1997; Betz and Hackett 1981). The gender difference in current class self-efficacy was evident immediately, prior to receiving substantial class feedback. Before any achievement feedback was provided, women expressed lower self-efficacy toward their current physics class than men, while expressing comparable self-efficacy beliefs toward other domains. After receiving class feedback, measured by the test average of the first two examinations, this gap reduced slightly but the self-efficacy of women toward their current physics class was still significantly lower than the current class self-efficacy of men mid-semester. This observation was unexpected and is not understood; more

research is required to understand this immediate difference in self-efficacy for women upon entering a STEM class.

While the main effect of time was not significant, current class self-efficacy did evolve during a class through a significant time-by-TestAve interaction. While statistically significant, the changes represented a small practical effect with a one-standard-deviation increase in TestAve producing only a 6.0% increase in current class self-efficacy. The gender-by-time-by-TestAve interaction was not significant suggesting that men and women processed test information into their current class self-efficacy in the same way. The gender-by-time interaction was significant in one of the two physics classes showing the gap in current class self-efficacy, which was present early in the class, narrowed somewhat by mid-semester in that class. The parallel evolution of self-efficacy with time while maintaining the same mean difference has also been observed in chemistry courses (Villafañe et al. 2014).

Self-Efficacy and Gender The role of gender in the model was not as hypothesized. Figure 2 shows that except for the initial difference in current class self-efficacy at the beginning of the class, gender does not affect the evolution of self-efficacy within the class. This is counter to a substantial strand of research showing that men and women incorporate mastery experiences into self-efficacy differently (Marshman et al. 2018; Sawtelle et al. 2012a, b; Vermeer et al. 2000; Zeldin et al. 2008; Zeldin and Pajares 2000). Achieving successful test results must be viewed as one of the primary mastery experiences of students in introductory physics classes. Neither the evolution of current class self-efficacy nor STEM self-efficacy as the students matriculated from Phys 1 to Phys 2 was moderated by gender.

In general, except for the difference in current class self-efficacy which were evident immediately upon entering the class, self-efficacy evolved in a similar fashion for both men

and women. There was no evidence that men and women processed course feedback in the form of test grades or class completion differently.

Implications and Future Work

The current study shows that, by the time students reach college, STEM self-efficacy changes little with additional academic feedback. This study provides the beginning of an analysis of how self-efficacy interacts with test performance feedback at the college level and how various forms of domain-specific self-efficacy change with time. Much additional research will be needed to fully understand the dynamic evolution and interaction of multiple domains of self-efficacy implied by the SCT.

The differences observed between current class self-efficacy and STEM self-efficacy for all students and the additional difference in current class self-efficacy observed for women imply that studies examining STEM self-efficacy may overestimate current class self-efficacy particularly for women. This effect could lead to important consequences if self-efficacy information is used for retention efforts. The identification of STEM self-efficacy as a stable construct may mean that efforts directed at improving STEM self-efficacy at the college level may meet with limited success.

The observation that the self-efficacy differences were present very early in the semester suggests that the differences did not result from the instruction in the classes studied, but instead were the result of factors outside the classes. This suggests that the instructional environment in the class studied was not producing the differences and does not need to be modified, but rather that the classes should institute policies that attempt to reduce already existing differences. These policies may include a number of intervention strategies which have been demonstrated to promote self-efficacy including providing mastery experiences, exposing students to successful peer role models, and providing positive goal-oriented formative feedback (Hazari et al. 2010; Schunk and Ertmer 2000).

Limitations

We acknowledge that test average was not the only influence on self-efficacy in each class; however, for all courses, test average made up the majority of the student's course grade. Other assignments in the courses such as laboratory activities and homework are often worked collaboratively, and therefore, the grades on these assignments are not a good individual measure of student achievement. This quantitative study was performed at a single institution in which the gender composition of the samples in many of the courses studied was quite

unbalanced. Additional qualitative and quantitative research at different institutions with differing demographic composition is needed to determine if results are similar in other contexts. This work was targeted toward introductory physics and mathematics courses; it should be extended to introductory chemistry courses to determine if the results are general for physical science courses.

Conclusion

This work showed that STEM students have three levels of academic STEM self-efficacy; the lowest toward the class they are currently in (current class self-efficacy), the next highest toward STEM classes in general (STEM self-efficacy), and the highest toward their intended profession (professional self-efficacy). Men and women reported having similar STEM self-efficacy and professional self-efficacy; however, there was a difference between current class self-efficacy, with men reporting higher current class self-efficacy. This gap was present very early in the class before substantive performance feedback was available. Except for this initial difference in self-efficacy, both men and women processed test information into their current class self-efficacy in the same way. General academic STEM self-efficacy changed little as students matriculated between classes and was not influenced by general academic achievement measures such as cumulative GPA; as such, general academic STEM self-efficacy may be a stable construct once students reach the second year of college.

Acknowledgements John Stewart and Rachel Henderson should be considered co-first authors of this manuscript.

Funding Information This work was partially supported by the National Science Foundation Education and Human Resources Core Research grant (ECR-1561517).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.

- Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Upper Saddle River: Prentice-Hall, Inc..
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Ed.), *Encyclopedia of Human Behavior* (Vol. 4, pp. 71–81). New York: Academic Press.
- Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: Freeman.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
- Besterfield-Sacre, M., Moreno, M., Shuman, L. J., & Atman, C. J. (2001). Gender and ethnicity differences in freshmen engineering student attitudes: a cross-institutional study. *Journal of Engineering Education*, 90(4), 477–489.
- Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28(5), 399–410.
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, 43(5), 485–499.
- Britner, S. L. (2008). Motivation in high school science students: a comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching*, 45(8), 955–970. <https://doi.org/10.1002/tea.20249>.
- Brown, S. D., Tramayne, S., Hoxha, D., Telander, K., Fan, X., & Lent, R. W. (2008). Social cognitive predictors of college students' academic performance and persistence: a meta-analytic path analysis. *Journal of Vocational Behavior*, 72(3), 298–308.
- Byars-Winston, A., Diestelmann, J., Savoy, J. N., & Hoyt, W. T. (2017). Unique effects and moderators of effects of sources on self-efficacy: a model-based meta-analysis. *Journal of Counseling Psychology*, 64(6), 645–658.
- Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: longitudinal study. *British Journal of Educational Psychology*, 81(1), 78–96.
- Cavallo, A. M., Potter, W. H., & Rozman, M. (2004). 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 Science and Mathematics*, 104(6), 288–300.
- Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review*, 76(5), 641–666.
- Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1), 1–35.
- Concannon, J. P., & Barrow, L. H. (2009). A cross-sectional study of engineering students' self-efficacy by gender, ethnicity, year, and transfer status. *Journal of Science Education and Technology*, 18(2), 163–172.
- Concannon, J. P., & Barrow, L. H. (2012). A reanalysis of engineering majors' self-efficacy beliefs. *Journal of Science Education and Technology*, 21(6), 742–753.
- Dalgety, J., & Coll, R. K. (2006). Exploring first-year science students' chemistry self-efficacy. *International Journal of Science and Mathematics Education*, 4(1), 97–116.
- Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40(2), 117–128.
- Hackett, G., Betz, N. E., Casas, J. M., & Rocha-Singh, I. A. (1992). Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology*, 39(4), 527–538.
- Hall, J. M., & Ponton, M. K. (2005). Mathematics self-efficacy of college freshman. *Journal of Developmental Education*, 28(3), 26–32.
- Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: a gender study. *Journal of Research in Science Teaching*, 47(8), 978–1003.
- Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: a new conceptualization of an established questionnaire. *SAGE Open*, 3(4), 2158244013510305.
- Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education*, 28(1), 1–35.
- Hutchison, M. A., Follman, D. K., Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, 95(1), 39–47.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527.
- Jagacinski, C. M. (2013). Women engineering students: competence perceptions and achievement goals in the freshman engineering course. *Sex Roles*, 69(11–12), 644–657.
- Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Unpacking gender differences in students' perceived experiences in introductory physics. In *AIP Conference Proceedings*, 1179(1), 177–180. College Park: AIP Publishing.
- Larose, S., Ratelle, C. F., Guay, F., Senécal, C., & Harvey, M. (2006). Trajectories of science self-efficacy beliefs during the college transition and academic and vocational adjustment in science and technology programs. *Educational Research and Evaluation*, 12(4), 373–393.
- Larson, L. M., Pesch, K. M., Surapaneni, S., Bonitz, V. S., Wu, T. F., & Werbel, J. D. (2014). Predicting graduation: the role of mathematics/science self-efficacy. *Journal of Career Assessment*, 23(3), 399–409.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, 33(3), 265.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1987). Comparison of three theoretically derived variables in predicting career and academic behavior: self-efficacy, interest congruence, and consequence thinking. *Journal of Counseling Psychology*, 34(3), 293.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45(1), 79–122.
- Lent, R. W., Sheu, H. B., Singley, D., Schmidt, J. A., Schmidt, L. C., & Gloster, C. S. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior*, 73(2), 328–335.
- Lindström, C., & Sharma, M. D. (2011). Self-efficacy of first year university physics students: do gender and prior formal instruction in physics matter?. *International Journal of Innovation in Science and Mathematics Education*, 19(2), 1–19.
- Mamaril, N. A., Usher, E. L., Li, C. R., Economy, D. R., & Kennedy, M. S. (2016). Measuring undergraduate students' engineering self-efficacy: a validation study. *Journal of Engineering Education*, 105(2), 366–395.
- Marra, R. M., Rodgers, K. A., Shen, D., & Bogue, B. (2009). Women engineering students and self-efficacy: a multi-year, multi-institution study of women engineering student self-efficacy. *Journal of Engineering Education*, 98(1), 27–38.
- Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: a

- cause for alarm? *Physical Review Physics Education Research*, 14(2), 020123.
- Nissen, J. M. (2019). Gender differences in self-efficacy states in high school physics. *Physical Review Physics Education Research*, 15(1), 013102.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24(2), 124–139.
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: the need for specificity of assessments. *Journal of Counseling Psychology*, 42(2), 190–198.
- Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). *A Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801–813.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387.
- Sawtelle, V., Brewé, E., Goertzen, R. M., & Kramer, L. H. (2012a). Identifying events that impact self-efficacy in physics learning. *Physical Review Physics Education Research*, 8(2), 020111.
- Sawtelle, V., Brewé, E., & Kramer, L. H. (2012b). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching*, 49(9), 1096–1121.
- Schunk, D. H., & Ertmer, P. (2000). Self-regulation and academic learning: self-efficacy enhancing interventions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-regulation* (pp. 631–649). NY: Academic Press.
- Schunk, D. H., & Pajares, F. (2007). The development of academic self-efficacy. In A. Wigfield & J. Eccles (Eds.), *Development of Achievement Motivation*. Hoboken, NJ: Wiley.
- Shaw, K. A. (2004). The development of a physics self-efficacy instrument for use in the introductory classroom. In *Physics Education Research Conference Proceedings*, 720(1), 137–140. College Park: AIP Publishing.
- Uitto, A. (2014). Interest, attitudes and self-efficacy beliefs explaining upper-secondary school students' orientation toward biology-related careers. *International Journal of Science and Mathematics Education*, 12(6), 1425–1444.
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796.
- Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, 92(2), 308–315.
- Villafañe, S. M., Garcia, C. A., & Lewis, J. E. (2014). Exploring diverse students' trends in chemistry self-efficacy throughout a semester of college-level preparatory chemistry. *Chemistry Education Research and Practice*, 15(2), 114–127.
- Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A social cognitive construct validation: determining women's and men's success in engineering programs. *The Journal of Higher Education*, 78(3), 337–364.
- Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching*, 45(9), 1036–1058.
- Zeldin, A. L., & Pajares, F. (2000). Against the odds: self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal*, 37, 215–246.
- Zimmerman, B. J. (2000). Self-efficacy: an essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.