

## Gender, experience, and self-efficacy in introductory physics

Jayson M. Nissen<sup>1</sup> and Jonathan T. Shemwell<sup>2</sup>

<sup>1</sup>*Department of Physics & Astronomy, University of Maine, Orono, Maine 04469, USA*

<sup>2</sup>*College of Education & Human Development, University of Maine, Orono, Maine 04469, USA*

(Received 6 February 2015; published 1 August 2016)

[This paper is part of the Focused Collection on Gender in Physics.] There is growing evidence of persistent gender achievement gaps in university physics instruction, not only for learning physics content, but also for developing productive attitudes and beliefs about learning physics. These gaps occur in both traditional and interactive-engagement (IE) styles of physics instruction. We investigated one gender gap in the area of attitudes and beliefs. This was men's and women's physics self-efficacy, which comprises students' thoughts and feelings about their capabilities to succeed as learners in physics. According to extant research using pre- and post-course surveys, the self-efficacy of both men and women tends to be reduced after taking traditional and IE physics courses. Moreover, self-efficacy is reduced further for women than for men. However, it remains unclear from these studies whether this gender difference is caused by physics instruction. It may be, for instance, that the greater reduction of women's self-efficacy in physics merely reflects a broader trend in university education that has little to do with physics *per se*. We investigated this and other alternative causes, using an in-the-moment measurement technique called the Experience Sampling Method (ESM). We used ESM to collect multiple samples of university students' feelings of self-efficacy during four types of activity for two one-week periods: (i) an introductory IE physics course, (ii) students' other introductory STEM courses, (iii) their non-STEM courses, and (iv) their activities outside of school. We found that women experienced the IE physics course with lower self-efficacy than men, but for the other three activity types, women's self-efficacy was not reliably different from men's. We therefore concluded that the experience of physics instruction in the IE physics course depressed women's self-efficacy. Using complementary measures showing the IE physics course to be similar to others in which gendered self-efficacy effects have been consistently observed, we further concluded that IE physics instruction in general is likely to be detrimental to women's self-efficacy. Consequently, there is a clear need to redress this inequity in IE physics, and probably also in traditional instruction.

DOI: [10.1103/PhysRevPhysEducRes.12.020105](https://doi.org/10.1103/PhysRevPhysEducRes.12.020105)

### I. INTRODUCTION

Over the last 60 years, physics has lagged behind other science, technology, engineering, and mathematics (STEM) disciplines in the proportion of women who pursue undergraduate degrees. For many STEM disciplines, the number of women relative to men is now at or near parity. As examples, between 2000 and 2010 women made up 50% of degree recipients in chemistry and 41% in mathematics. However, during this same period only 21% of bachelors degrees in physics were received by women [1].

One reason why so few women may be pursuing physics degrees is that the physics-learning environment preferentially favors male students over female students. This possibility is backed by research showing persistent differences in how women and men experience physics, in which women are disadvantaged. In introductory courses,

women tend to both start out and end at lower levels of conceptual knowledge than men [2,3]. Furthermore, women tend to have less productive attitudes about learning physics, including interest, sense making effort, and problem solving confidence [3]. For both conceptual knowledge and attitudes, these gender differences increase from pre- to post-course measurement [2,3].

The gender gap in attitudes and beliefs about physics learning also extends to self-efficacy, our subject here. Self-efficacy is the belief in one's ability to succeed in a given domain [4]. It is an important predictor of academic performance and persistence, both in general [5], and in introductory physics courses [6,7]. Kost-Smith [7] found that women entered introductory physics courses with lower self-efficacy than men, and this difference increased from pre- to post-course. Sawtelle *et al.* [8] obtained the same result in lecture-based physics courses, as did Cavallo *et al.* [9] and Lindstrom and Sharma [10].

While it seems fairly clear that there is a gender gap in self-efficacy in physics, it remains an open question whether physics instruction somehow causes this inequity. This is a very important question and the central one of the present

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

study. It may be, for instance, that the negative shift in women's self-efficacy that is consistently observed in physics is not unique to physics courses. Rather, this shift may be an epiphenomenon, or secondary effect, of a broader trend that would tend to occur in most courses, or perhaps most STEM courses. So long as this and other broad-based causes of the inequity (these are discussed later in this article) cannot be ruled out, then there is no particular urgency to redress it in physics courses. However, if gender differences in self-efficacy could be shown to be caused by physics instruction, then there would be an obvious need for concerted action within the physics community to bring about more equitable classroom experiences. The purpose of the present study is to see whether the possibility of an epiphenomenon and other explanations could be ruled out, thus resting the source of the gender inequity more squarely on physics instruction.

We engaged with the question of causality by measuring men's and women's feelings of self-efficacy as they were learning in physics and in other STEM and non-STEM courses over two weeks of instruction. The measurement used an established quantitative technique called the Experience Sampling Method (ESM) in which students responded to a signal to briefly record their thoughts and feelings of self-efficacy in the midst of their activities. We reasoned that if women could be observed to experience lower self-efficacy than men in physics, but not in other courses, physics instruction would have to be seen as a primary cause of the gender difference.

Bandura ([4], p. 3) defined self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." The term "beliefs" in this definition is potentially confusing because it suggests that self-efficacy is a fairly stable characteristic. However, Bandura considered self-efficacy to be "a dynamic fluctuating property, not a static trait" ([4], p. 406). Furthermore, he recognized that it was highly responsive to a person's behavior and their environment. Therefore, self-efficacy beliefs are sometimes better thought of as being dynamic states. On the other hand, Bandura acknowledged that self-efficacy was often associated with habitual patterns of behavior [4,11]. Accordingly, self-efficacy is sometimes measured using surveys that ask people to rate their confidence in their ability to accomplish tasks, with the results interpreted as being traitlike characteristics [3,11]. Thus, a person's self-efficacy in physics can be said to go up or down after a semester of instruction.

To encompass both dynamic and stable aspects of self-efficacy, hereafter we refer to (and measure) them as two distinct components. One is the dynamic response that may shift from moment to moment, which we call the self-efficacy state. The second is a more stable attitude (or belief) about one's ability to succeed in a domain, which we refer to as the self-efficacy trait.

So far as we can tell, our approach to measuring self-efficacy states separately from traits is unique. Often,

researchers skirt the issue by making the sources of self-efficacy the object of measurement, rather than self-efficacy itself. Typically, this is done by asking people to rate their agreement with statements about experiences they had in the domain of interest [8,12]. Since the sources of self-efficacy are assumed to underlie both dynamic states and longer term patterns, there is no need to distinguish between these aspects. In physics, the Sources of Self-Efficacy in Science Courses-Physics (SOSESC-P) [12] takes this approach, asking students to reflect on sources of self-efficacy in their experiences of physics instruction. Another instrument used in physics asks students about their self-efficacy via their confidence in their ability to succeed at physics tasks, obtaining a measurement of self-efficacy beliefs, which we refer to as self-efficacy traits. This is the Physics Self-Efficacy and Identity Survey (PSEIS) [7]. This instrument also includes sources of self-efficacy items from the SOSESC-P.

### A. The gender gap in physics self-efficacy

Leaders in the field have pointed out that the development of coherent attitudes and beliefs about learning and doing science should be a core goal of physics education [13,14]. Unfortunately, these attitudes and beliefs generally erode over time in physics courses, even when using research-based pedagogies that manifestly benefit learning [3,14,15]. Furthermore, there are consistent gender differences in which negative shifts in attitudes and beliefs are larger for women than for men [3,15]. These differences extend to self-efficacy. Using the PSEIS, Kost-Smith [7] demonstrated that women had larger negative shifts than men for both self-efficacy traits and sources of self-efficacy. This result was reliable across four instructors and three different offerings of a research-based introductory course known as interactive-engagement physics. Kost-Smith also found that there was a gender difference in conceptual knowledge at post-test in these courses (typical for physics instruction), and that 12% of the effect was predicted by gender differences in self-efficacy beliefs. Sawtelle *et al.* [8] used the SOSESC-P to show that students' physics self-efficacy became less positive across three different semesters of lecture-based physics courses, with the negative shift consistently larger for female students. Corroborating evidence for a reliable gender gap in self-efficacy beliefs in physics, at least in introductory courses, comes from studies of general attitudes and beliefs in physics. Most notably, Kost *et al.* [3] and Kost-Smith *et al.* [15] found that women started interactive engagement physics courses with less expert attitudes about learning and doing physics than men, and these differences tended to increase from pre- to postinstruction.

Sawtelle *et al.* [8] pointed out a notable exception to the trend in negative and gendered self-efficacy outcomes in physics. Studying a course that used modeling instruction, they measured self-efficacy traits at the beginning and end

of the course for three different semesters using the SOSESC-P. They found neither positive nor negative shifts in either men's or women's self-efficacy traits. Sawtelle *et al.* [16] investigated the source of this salutary outcome using video and interviews of three students engaged in modeling activities. They showed that creating and working on models in small group settings (the primary instructional mode of the course) provided many opportunities for self-efficacy development, such as when students received positive feedback from their classmates and vicariously from seeing their classmates succeed. They proposed that these opportunities might be what differentiated modeling instruction from other physics courses with regard to self-efficacy outcomes.

Assuming that modeling instruction in general does not negatively affect women's physics self-efficacy, then the gendered self-efficacy outcomes found in traditional and IE physics would be more likely to be caused by the experience of instruction in those formats and less likely to result from a persistent, broad-based trend in university education. However, Sawtelle and colleagues' [8,16] research was not intended to be conclusive about the causes of self-efficacy outcomes in modeling instruction. Sawtelle and colleagues' second, more diagnostic study in particular was not intended to explain variance but rather to reveal processes by which self-efficacy could be supported. Thus, direct evidence of the impact of more mainstream (i.e., non-modeling) physics instruction on men's and women's self-efficacy is needed if its gender effects are to be squarely established.

## B. Classroom environments, experiences, and gender

Much of the more general education research on differences in how male and female students experience STEM instruction has focused on the tenor of the classroom set by the professor. Using interviews, Hall and Sandler [17] found that women experienced "chilly" classrooms in which male instructors maintained classroom inequalities such as spending disproportionate amounts of time talking to male students and ignoring female students' questions. Seymour and Hewitt [18] used interviews to show low levels of faculty support and highly competitive environments were typically the starting point of students' paths out of STEM majors. They concluded that many highly capable students, including women, were leaving STEM disciplines because of their poor experiences and not because of an inability to perform well in their coursework. In physics, Mujtaba and Reiss [19] analyzed high school students' end-of-course surveys to show a gender difference in the level of encouragement to continue in the discipline they felt from their teachers. This measure was correlated with students' intentions to take additional physics courses in the future. Similarly, Kost *et al.* [20] used a survey to show that women reported experiencing less support in physics courses, for instance, compared to

men, women more frequently agreed with the item "I felt like I didn't belong in this course."

The gender inequities just described are relevant to the present study because they are attributed to the experience of learning rather than within a broader gender-based trend. However, these studies used retrospective measures, wherein the distance from the experience of instruction leaves open the possibility of alternative causal factors. In particular, gender differences could arise because men and women focus on different aspects of their experiences in retrospection; not because they actually experienced instruction differently. For example, Hyde *et al.* [21] found that women retrospectively reported greater levels of anxiety about mathematics than men. They inferred from this result that women experienced higher levels of anxiety during their mathematics courses. Goetz *et al.* [22] called this interpretation into question by combining retrospective reports with an in-the-moment measure of anxiety, which was the Experience Sampling Method used in the present study. Retrospective surveys found that women reported higher levels of mathematics anxiety than men, but the in-the-moment measure showed that women and men experienced very similar levels of anxiety. Bieg *et al.* [23] referred to this mismatch as a state-trait discrepancy. They found that much of it was explained by students' mathematics self-concept, which they described as a measure of students' feelings of control over their performance in the course. They proposed that the state-trait discrepancy arose when students with lower math self-concepts focused more on their anxiety in retrospective reporting than did students with higher math self-concept.

## II. THEORETICAL FRAMEWORK

As we discussed earlier there are inconsistencies in how self-efficacy is described as both dynamic and static, which we have addressed by separating self-efficacy into states and traits. Albert Bandura [4,24] proposed that internal states are one of the three major classes of determinants in human agency, along with behavior and environment. States arise within the individual, have a complex latent structure consisting of affect, cognition, and biological events and are dynamically responsive to both the perceived environment and the individual's behavior (see Fig. 1) [4]. In contrast, traits are the relatively stable patterns of behaviors and internal states, including thoughts and feelings, that habitually occur in different circumstances and contexts [25]. We propose thinking of traits as representing the patterns that arise between the three major classes of determinants: internal states, environment, and behavior. This framework is consistent with the definitions both of traits and of self-efficacy in that self-efficacy traits are context and situation dependent, tend to be very stable, and result in habitual patterns of behavior [4,11].

The development of self-efficacy traits is rooted in experience [26]. High levels of performance support the



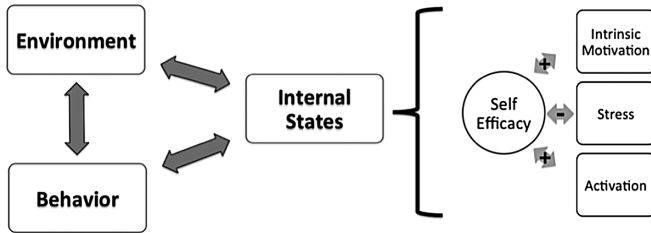


FIG. 1. The three major classes of determinants according to Social Cognitive Theory are shown on the left. The arrows represent the reciprocal causal relationships that exist between each of the classes. On the right, the internal state class is broken down into the four affective states that were measured in this study. The relationships between the self-efficacy state and each of the complementary states are shown in the arrows.

development of stronger self-efficacy traits which subsequently support future performance [27]. Because self-efficacy states are a measure of experience and, to some degree, a measure of personal performance, we expected a similar causal reciprocal relationship to exist between self-efficacy states and traits measured in the present study. Therefore, self-efficacy states experienced in physics integrated over time should produce physics self-efficacy traits. Therefore, because self-efficacy traits predict student performance in physics [6,7], we viewed very low self-efficacy states experienced in physics as harmful to students' persistence and success in physics.

### III. RESEARCH QUESTIONS

The negative shift in women's physics self-efficacy traits measured across introductory physics instruction [7–10] suggests that there is something about physics instruction that is particularly harmful to women's self-efficacy compared to men's. However, as we discussed earlier, rival explanations that this is caused by factors outside of the experience of instruction must be dealt with before locating the cause within physics instruction and not elsewhere. The main body of research to date has primarily focused on post-course measures and/or only on physics courses, so it has not effectively addressed these rival possibilities. In order to address the overarching question about the cause of the larger negative effect on women's self-efficacy being situated in the experience of physics instruction we asked two principle research questions:

- (1) To what extent did women experience IE physics instruction with lower self-efficacy states than men?
- (2) How did the differences between men's and women's self-efficacy states in IE physics compare to the differences in other STEM and non-STEM courses?

### IV. METHODS

#### A. Context

The study took place at a four-year public university located in the northeastern part of the United States. The

university was the leading research university for the state it served and was a PhD-granting institution in many STEM fields.

We collected data in one interactive engagement (IE) physics course, the focal IE physics course. Interactive engagement promotes [28] [p. 65] “conceptual understanding through interactive engagement of students in heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors.” IE has been used to describe courses that use research based teaching practices [3,29,30] such as Peer Instruction [31] and Tutorials in Introductory Physics [32].

We collected data in an IE physics course, as opposed to a traditional physics course, because we expected IE instruction to be a more conservative measure of gender differences in self-efficacy states. We based this decision on IE instruction better supporting student conceptual learning, and because gender differences in conceptual knowledge tend to be smaller after IE instruction than after traditional physics instruction [2].

The focal IE physics course met five times in total each week: twice for 50 min of lecture with approximately 150 students, twice for 50 min of recitation with 24 students and once for 110 min of laboratory with 24 students. The instructor of the course was male and had 35 years of teaching experience. The data were collected during the fifth year the instructor taught this course. The course was modeled on IE physics courses described in Kost *et al.* [3]. Almost all lectures used several conceptual multiple choice questions embedded throughout the lecture, i.e., ConcepTests [31]. Students discussed these questions with their neighbors and the course instructor called on students to explain their reasoning for their answers. Students earned a small portion of their final course grade, 3%, by participating in the ConcepTests. Three midterm exams and one final exam were given in the lecture portion of the course. There was a weekly homework assignment with a written and an online component. Homework and tests included both conceptual and calculation problems. In the two recitation sections students spent most of their time solving conceptual problems in small groups. One recitation per week made use of a standard set of tutorial lessons [32]. The other recitation used a mix of locally generated conceptual and calculation physics problems. A graduate teaching assistant (TA) facilitated the recitation periods and the lab. An undergraduate learning assistant (LA) assisted the TA during the recitation. The LA had previously completed the course and was enrolled in a weekly seminar on pedagogy [33]. The TA and LA were provided weekly training on the content and pedagogy used during recitation. This training emphasized the use of Socratic dialogue to support students in generating their own conceptual understanding in the activities during recitation.

**B. Design**

The study used a within-subject design comparing students' self-efficacy states in the focal IE physics course, to their self-efficacy states in other introductory STEM courses the students were taking in the same semester (see Fig. 2). This design enabled us to address five research goals to provide evidence for answering the research questions (goals 1 and 2), provide validity for that evidence (goals 3 and 4), and generalize the findings (goal 5). The first three goals were addressed with the state data; the last two goals were addressed with the trait data.

For goal 1 we identified any gender differences in the self-efficacy states students experienced during instruction in the focal IE physics course and the size of those differences. For goal 2 we determined the extent to which any gender differences in self-efficacy states were unique to the focal IE physics course or whether they also occurred in the other courses, potentially as part of a broader trend, by comparing self-efficacy states in the focal IE physics course to those in other STEM courses (Fig. 2, left). For goal 3 we determined the extent to which gender differences in the complementary states were consistent with gender differences in the self-efficacy states experienced in the focal IE physics course.

An important feature of the design was the use of complementary measures to provide validity for any

identified gender difference in self-efficacy states that were measured in the focal IE physics course, goals 3 and 4. In addressing goal 4 self-efficacy trait data were used to validate the self-efficacy state data. The self-efficacy trait measure complemented the self-efficacy state measure, as shown by the dark arrow in Fig. 2, in that gender differences in self-efficacy states experienced in the IE physics course should show up as gender differences in the means of, and shifts in, self-efficacy traits across the semester. A secondary objective of goal 4 was the use of complementary trait measures, bottom right of Fig. 2, to support the validity of the self-efficacy trait measure by identifying the extent to which gender differences were consistent across all traits.

Because we studied only one semester of a single IE physics course, we designed the research to collect evidence of how well this focal course represented IE physics courses in general, goal 5. We compared the scores and gender differences in scores to those scores for similarly designed courses at another institution as reported by Kost *et al.* [3]. To do this, we used three different pre-post measures relying on standard survey instruments: self-efficacy traits [7], attitudes [34], and conceptual knowledge [35]. A fourth comparative measure was course grades. In Fig. 2, the latter three measures are grouped at the lower right under the collective heading of complementary traits.

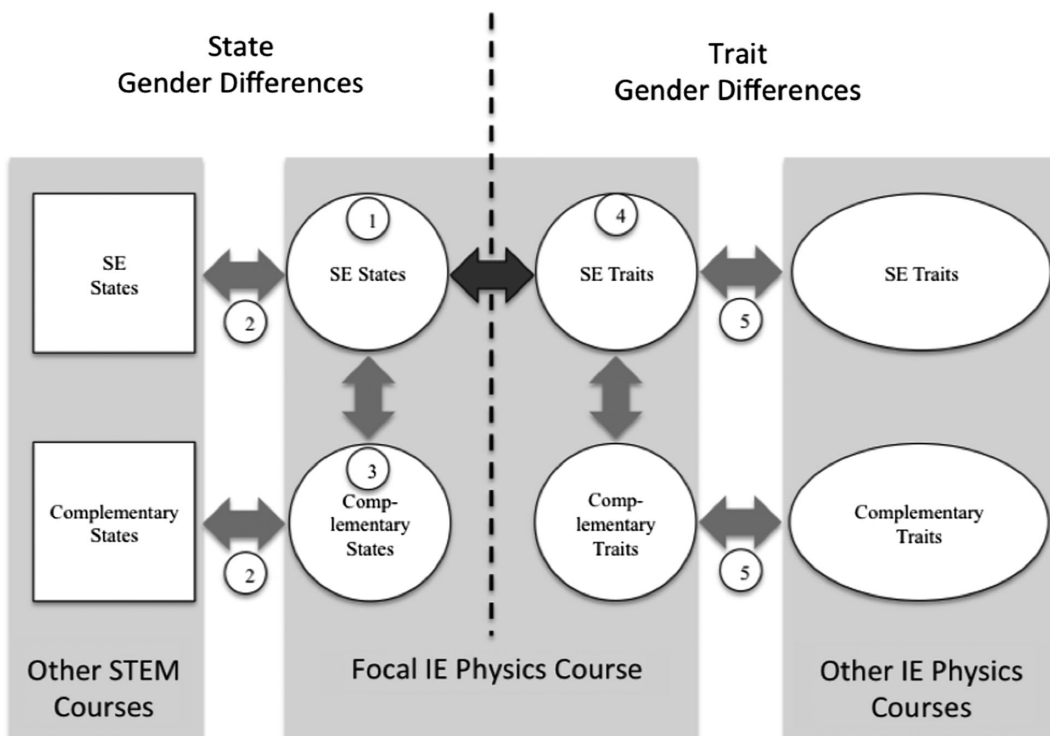


FIG. 2. Design structure of the research illustrating the five goals of the design. (1) Identify gender differences in self-efficacy states experienced in the IE physics course, (2) identify if the gender differences in state experiences were unique to IE physics, (3) consistency of gender differences in the focal IE physics course for self-efficacy states and the complementary states, (4) consistency between gender differences for self-efficacy states and traits, and (5) identify how similar trait outcomes and gender differences were in the focal IE physics course to courses studied by Kost *et al.* [3].

Because of the intensive nature of the ESM, it is typical to collect state data from a representative sample of participants in a given context, such as a course or a school, rather than taking data from all students. Using this approach, we conducted ESM with 33 *ESM participants* from a physics course of 242 students. By contrast, trait data were much easier to collect and we obtained this from a larger sample of 117 *trait participants*. Unfortunately, not all 33 ESM participants were part of the 117 trait participants (see Fig. 3). Therefore, the two overlapping groups were used as independent samples for different purposes. Trait participants were used to represent the effect that the course had on students' traits. ESM participants were used to characterize how students experienced instruction.

### C. Participants

Out of 242 students who started the course, 222 completed the course and received grades. Of these, 40 (18%) were female (see Fig. 3). Of the 20 students who dropped or withdrew from the course, 5 were female. Of the 117 trait participants 90 were male and 27 (23%) female. Of the 33 ESM participants 20 were male and 13 (39%) were female. Overall, there were 20 ESM participants who were also trait participants, 12 male and 8 female. Two of the female ESM participants withdrew from the course and did not receive final course grades.

ESM participants were recruited from their IE physics course through a brief announcement by the first author describing the research. They were informed that the research was investigating their experiences as college

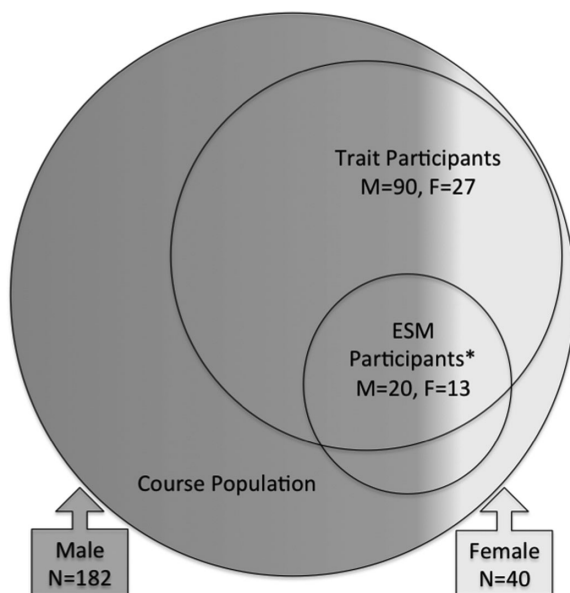


FIG. 3. Diagram of the overlapping ESM participants and trait participants. \* Eight male and five female ESM participants were not trait participants.

students. All students who wished to participate in the study were allowed to do so. Participants who completed the ESM were given a small amount of extra credit and a stipend of fifty USD.

We defined gender as the self-identification as either male or female.

### D. Instrumentation for trait data collection

We measured students' *self-efficacy traits* in physics by using the 20 5-point Likert-scale self-efficacy questions from the Physics Self-Efficacy and Identity Survey developed by Kost-Smith [7]. We truncated the name to Physics Self-Efficacy Survey (PSES) because we did not include the identity questions or the sources of self-efficacy questions. The PSES measures self-efficacy across four constructs, but only the overall self-efficacy score was used in this study. We measured students' *attitudes* about learning physics with the Colorado Learning Attitudes about Science Survey (CLASS) [34]. The CLASS measures eight separate categories of student beliefs compiled from student responses to 42 questions. Responses are coded as favorable, neutral, or unfavorable based on agreement with expert responses. Like the PSES, the CLASS is multidimensional, having eight subconstructs of expertlike response, but it also allows for an aggregate score. We used only the overall favorable score in the present study. We measured students' *conceptual knowledge* in the focal IE physics course with the Force and Motion Conceptual Evaluation (FMCE) [35], a 47 question multiple-choice exam. The FMCE was scored out of 37 points following the methods of Thornton *et al.* [36], using a spreadsheet developed for that purpose [37]. We obtained *course grades* for the focal IE physics course from the instructor and analyzed them on a 4.0 scale, such that an A was 4.0, an A- was 3.7, a B+ was 3.3, etc. This was the scale used at this institution and was the same scale used by Kost *et al.* [3].

### E. Experience Sampling Form

The data collection instrument for ESM studies is a short survey that participants fill out when randomly signaled, or shortly thereafter, about the activity they were engaged in at the moment of the signal. ESM studies typically refer to this instrument as the Experience Sampling Form (ESF). Our ESF was modeled on those used in studies overviewed by Hektner *et al.* [38]. It was the single side of one standard-sized page split into two sections. The first section asked four free-response questions: (1) the main and (2) the secondary activities students were doing, (3) where they were, and (4) what they were thinking about. For the present study, only the first of these free-response items was analyzed. The second portion of the survey, on the right half of the page, consisted of 20 Likert-scale questions. Students indicated the type and level of affect at the moment they were signaled by responding to the question,

TABLE I. Affective state constructs, definitions and component questions with construct reliability measures and factor loadings. Italicized questions were asked in a 7-point bipolar format. All other questions were asked in a 5-point unipolar format. Numbers in the left hand column are: Cronbach's alpha (percent variance explained). Parentheses in the right hand column are the rotated factor loading for that question.

Construct	Definition	Components
Self-efficacy 0.76(20.2%)	Dynamically responsive judgments of one's ability to organize and execute the courses of action required to produce given attainments in the activity at hand.	skill(0.79), control(0.68), success(0.82), <i>difficulty concentrating easy to hard</i> (−0.51), <i>confused to clear</i> (0.52)
Activation 0.87(25.6%)	An elevated level of excitement and involvement in the task, consistent with Thayer [39] and in contrast to a relaxing state [40].	determined(0.67), active(0.59), attentive(0.80), alert(0.77), inspired(0.56), <i>detached to involved</i> (0.57), concentrating (0.63)
Intrinsic motivation 0.70(6.1%)	A drive to engage in the activity at hand, derived from within, either because it is personally enjoyable or valuable, as opposed to extrinsic motivation, which is driven by external pressures or rewards [41,42].	<i>free to constrained</i> (−0.65), <i>excited to bored</i> (−0.64), enjoy (0.57), importance (−0.71)
Stress 0.79(6.6%)	Negative feelings resulting from an individual's perception that they do not have the resources to cope with a perceived situation [43].	stress(0.83), worry(0.80), frustration(0.71)

“How did you feel in the main activity?” which was followed by the 20 emotions. Principle components analysis of all surveys confirmed that 19 of the 20 Likert-scale items reliably loaded onto the four affect constructs as shown in Table I.

Three of the Likert-scale questions, skill, control, and success, formed the basis of our self-efficacy state measure. We designed the study to include these questions because control and capability are central attributes of self-efficacy [4]. These feelings have also been statistically grouped in prior ESM studies [38,44] and principle components analysis confirmed their structure in the present study [45]. The other Likert-scale questions formed the three complementary affective states, activation, intrinsic motivation, and stress, which are defined in Table I. Self-efficacy, activation, and stress were measured on a unipolar scale from none to extreme. Intrinsic motivation was measured on a bipolar scale from extremely extrinsic to extremely intrinsic [46].

We used the relationships between self-efficacy and each of the complementary state measures to provide additional validity for the self-efficacy state measure. The relationship between self-efficacy and stress was expected to be negative. When self-efficacy is higher, stress should be lower because self-efficacy is a measure of personal skill and stress arises when skill does not meet the demands of the situation. The relationship between self-efficacy and both activation and intrinsic motivation was expected to be positive. People are more likely to become activated when they feel efficacious [4] and people are also more likely to internalize motivation for activities in which they feel efficacious [47].

## F. Procedures

ESM data were collected for two different seven-day periods during the third week and then again in the tenth

week of the semester. These weeks were chosen so as not to fall during an exam or other significant assessment. Signals to fill out the ESF were sent to students' cell phones. These were semirandomly scheduled across each day such that there was a signal once during each 2 h block between 8 a.m. and 10 p.m. and all signals were greater than 30 min apart. A constraint on the schedule was that a signal was scheduled for every physics course meeting, resulting in a higher rate of sampling for physics than for other experiences. We did this to ensure enough samples in physics for a reliable measurement, since students spent less time there than in the other, broader categories of experience. To prepare participants for the first of the two data collection periods, we gave them a 1 h briefing on the data collection procedures.

Surveys for trait measurements were given during the first and last week of the course. The knowledge measurement (FMCE) was done during class. This was not part of student grades but was a mandatory class activity for students in attendance. Students took the attitude and self-efficacy surveys (CLASS and PSES) outside of class via an online platform as a part of weekly homework assignments. Students received credit equal to one homework problem for completing each survey. We obtained course grades from the instructor after the course had ended.

## G. Methods of analysis

ESM data were analyzed to compare means between genders across all four activities for each of the four affective constructs. To check for statistical significance, we used a three-step process beginning with an omnibus multivariate analysis of variance to see if a statistically significant difference in means existed for the gender  $X$  activity interaction. Then factorial univariate analysis identified if statistically significant differences existed for the gender  $X$  activity interaction on each of the affective



constructs. Last, *post hoc* tests were run to identify statistically significant differences in means between males and females for each of the four affective constructs in each of the four activities.

The ESM data for both the third and tenth weeks of the semester were entered into a spreadsheet database. Principle components analysis was conducted on the raw responses and verified the individual questions aligned with the four expected affective constructs, summarized in Table I. Averaging the component questions of each construct on a 5-point, 0–4 scale created the raw score for each construct. The data for what students were doing were reduced to four activities: nonschool, non-STEM, STEM, and IE physics and the two weeks of data collection were combined. Analysis of variance confirmed that no statistically significant differences existed for either of these reductions. Results of these analyses are reported in Nissen [45].

We used Cohen's  $d$ , histograms of the raw score responses and  $Z$  scores of the affective constructs to interpret the size of the gender differences measured in the focal IE physics course. The histograms allowed comparing the distribution of students' responses across the scale for each affective construct. This supported interpreting the meaningfulness of the differences, for instance, in the case where one population never experienced a very high level of a state but the other population frequently experienced that high level.

$Z$  scores allowed identifying how the experience in physics was situated in students' overall experiences in two ways. First, they allowed showing how males' and females' average experiences in physics compared to their overall experiences, for example, bottom 20% or top 10%. Second, they allowed seeing how often physics experiences were above average. To create  $Z$  scores the 20 Likert-scale affect questions were converted to  $Z$  scores for each response based on that participant's mean and standard deviation for that question for that week. This conversion minimized the effects of participants using the scales differently by describing responses as above or below average for that person and scaling the distance from average in units of standard deviation for that person's response to that question [38]. Averaging the component question  $Z$  scores created the  $Z$  score for each affective construct.

Each of the four trait measures yielded a single overall score. We compared means for these scores between male and female students for all of the trait participants (i.e., all of the students for whom we had a complete set of trait data, see Fig. 2). We assessed the effect size of any differences between men's and women's mean scores on each trait measure using Cohen's  $d$ . To check for statistical significance, we used a two-step process beginning with an omnibus multivariate analysis of variance to see if statistically significant differences in means existed for

gender. Then, factorial univariate analysis identified if statistically significant gender differences existed on each measure.

We used results from the trait analysis to assess the similarity of the focal physics course to those investigated by Kost *et al.* [3] by comparing means for male and female students on each measure between the two course contexts. In particular, we compared the effect sizes for gender differences to see if the focal course maintained, increased, or decreased gender differences in similar ways to other IE physics courses [3].

Representativeness of trait participants was investigated by comparing mean grades of trait participants to mean grades for all other students while controlling for gender using analysis of variance. Assessing the representativeness of ESM participants was more challenging because we sought to balance the number of students included in the analysis with the number of trait measures over which we analyzed the representativeness. First, analysis of variance was used to compare means on all trait measures between ESM participants and nonparticipants. However, this limited the ESM participants included in the analysis in a biased way and the small  $N$  resulted in low statistical power. Therefore, analysis for representativeness of the ESM participants was accomplished by comparing means for ESM participants and nonparticipants on each trait measure for all students who completed that measure using two tailed  $t$  tests.

Cohen's  $d$  was utilized as a measure of the effect size between male and female students for both traits and state experiences as recommended by Rodriguez *et al.* [48]. Cohen [49] provided guidelines of small (0.2), medium (0.5), and large (0.8) for interpreting effect sizes for interventions, but cautions that these are not hard and fast rules. Thus, we used these guidelines for interpreting effect sizes loosely and described the differences in experience as descriptively as possible in order to substantiate the size of those differences.

## V. RESULTS

In presenting the results we first present the state data and then the trait data. We begin the state results by describing how well ESM participants represented the course population. Next, we present the results for self-efficacy states experienced in the IE physics course compared to other types of courses and day-to-day experiences. This addresses the two research questions and the first three design goals. We begin the trait results by describing the representativeness of the trait participants. Then we present the trait results to address the fourth and fifth design goals: checking the extent to which self-efficacy states in the focal IE physics course were consistent with physics self-efficacy traits and assessing the degree to which the focal IE physics course should be taken as representative of IE physics courses in general.



TABLE II. Representativeness of male and female ESM participants. Includes traits for ESM participants and nonparticipants by gender for all students who completed each trait instrument.

Measure	Scale	Participants			Nonparticipants			<i>d</i> (95% CI)
		Mean	<i>N</i>	SD	Mean	<i>N</i>	SD	
<i>Male</i>								
Pre-FMCE	%	32.9	17	17.5	29.7	163	23.3	0.14 (−0.36, 0.64)
Post-FMCE	%	69.1	16	26.0	63.2	145	28.7	0.20 (−0.31, 0.72)
Pre-CLASS	%	72.0	19	13.4	62.0	164	16.1	0.64 (0.16, 1.11)
Post-CLASS	%	64.8	16	13.8	58.3	101	17.5	0.38 (−0.15, 0.91)
Pre-PSES	1–5	3.61	19	0.41	3.43	156	0.59	0.31 (−0.17, 0.79)
Post-PSES	1–5	3.62	17	0.49	3.41	97	0.73	0.30 (−0.22, 0.81)
Course grade	0–4	3.05	20	1.11	2.31	162	1.31	0.57 (0.10, 1.04)
<i>Female</i>								
Pre-FMCE	%	22.7	13	10.3	23.4	31	18.2	−0.04 (−0.69, 0.60)
Post-FMCE	%	55.6	12	27.8	48.1	28	27.5	0.27 (−0.41, 0.95)
Pre-CLASS	%	55.8	13	19.4	61.8	28	16.6	−0.34 (−0.99, 0.33)
Post-CLASS	%	46.5	11	14.4	58.5	22	19.0	−0.68 (−1.41, 0.08)
Pre-PSES	1–5	3.23	11	0.62	3.34	29	0.58	−0.19 (−0.88, 0.51)
Post-PSES	1–5	3.14	10	0.40	3.10	23	0.55	0.08 (−0.67, 0.82)
Course grade	0–4	3.05	11	0.74	2.34	29	1.35	0.58 (−0.13, 1.28)

### A. The representativeness of ESM participants

Differences in traits between students who participated in the ESM and those who did not were tested by comparing the means between ESM participants and nonparticipants for male and female students for all students who completed each trait measure as shown in Table II. These comparisons showed that, first of all, both the male and female ESM participants in the study were high achieving students in the sense that they learned more conceptual knowledge and had higher grades than other students in the course. While there were no differences in selectivity between men and women with respect to achievement, the other trait measures suggested that men who participated in the ESM might have had especially robust attitudes and self-efficacy traits compared to the other men in the course. Whereas female ESM participants had more novicelike and more malleable attitudes, but similar self-efficacy traits, compared to other women in the course. These differences, or biases, provide a caveat for generalizing gender differences in the sample to the course population.

### B. Gender differences in self-efficacy and complementary states

The largest gender difference for self-efficacy states occurred in the focal IE physics course where women experienced much lower average self-efficacy (1.57) than men (2.25). There was a much smaller gender difference for mean self-efficacy states in other STEM courses with women having slightly lower means (2.25) than men (2.45). Thus, women experienced the focal IE physics

course with much lower self-efficacy than the other STEM courses, whereas the difference was relatively small for men. The second largest gender difference in experience was for intrinsic motivation in the focal IE physics course. Women experienced more extrinsic motivation (1.25) than men (1.61). There was a much smaller gender difference in other STEM courses with women having more extrinsic motivation (1.47) than men (1.64). Similar to self-efficacy, the difference between men's motivation in the focal IE physics course and in other STEM courses was small, whereas women's motivation was much more extrinsic in the focal IE physics course. Consistent with women's lower mean self-efficacy states and more extrinsic-motivation states in the focal IE physics course women also experienced greater stress in physics (1.48) than men (1.30) and lower activation (1.99) than men (2.13).

Analysis of variance was used to determine if any statistically significant differences existed between male and female students' experiences. A 2X4 MANOVA with independent variables for activity and gender and dependent variables for the four affective constructs identified statistically significant effects for gender, activity, and for the gender X activity interaction, Table III. The statistical significance of the gender X activity interaction indicated that there might have been statistically significant differences in experience for male and female students for some of the affective constructs specific to certain activities. This was tested with univariate analysis of variance and was statistically significant on the gender X activity interaction for self-efficacy and activation. The analysis for the activity condition is discussed elsewhere [45,46].

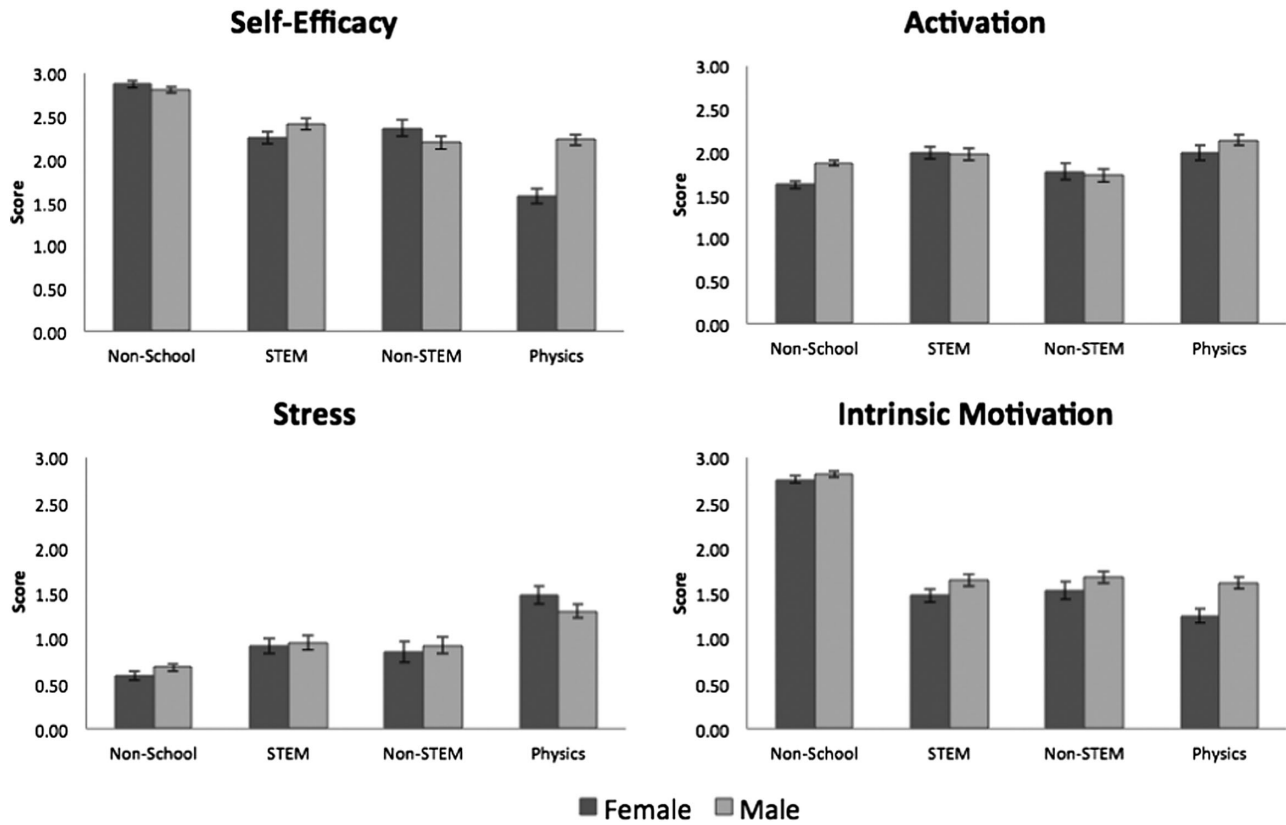


FIG. 4. Students’ affective state experiences by gender and activity. States were measured on a 5-point Likert scale and ranged from 0, not-at-all, to 4, extremely, for self-efficacy, stress, and activation. Intrinsic motivation ranged from extremely extrinsic, 0, to extremely intrinsic, 4. Compared to men, women in IE physics experienced lower self-efficacy, more extrinsic motivation, lower activation, and higher stress. Analysis indicated that the large gender differences for self-efficacy states were unique to the focal IE physics course. Error bars are 1 standard error.

Post hoc analysis further investigated the statistical significance of gender differences for each affective construct in each activity using two tailed *t* tests. Only two gender differences were statistically significant outside of the focal IE physics course, activation in nonschool and motivation in STEM courses, Table IV. In addition to the large gender differences for self-efficacy states in the focal IE physics course being statistically significant, so was the moderately large difference for intrinsic motivation. The small difference for activation was marginally statistically significant and the small difference for stress was not

statistically significant. These results portray a consistent picture of the focal IE physics course having been experienced more negatively by women, with the largest gender difference measured for self-efficacy states. In no other activities were there large or consistent gender differences in experience.

Z-score transformed data illustrated the size of the difference in men’s and women’s self-efficacy states experienced in the focal IE physics course. We accomplished this in two steps: first, by ranking their physics experiences within their overall experiences, and, second, by seeing how often

TABLE III. MANOVA and ANOVA results for affective state experiences.

	Gender		Activity		Gender X Activity	
	<i>F</i> (df, error df)	<i>p</i>	<i>F</i> (df, error df)	<i>p</i>	<i>F</i> (df, error df)	<i>p</i>
MANOVA	5.37 (4, 1429)	<0.001	335 (4, 1431)	<0.001	12.8 (4, 1431)	<0.001
<i>ANOVA Results</i>						
Self-efficacy	8.02 (1, 1432)	0.005	96.8 (3, 1432)	<0.001	13.4 (3, 1432)	<0.001
Activation	3.02 (1, 1432)	0.082	13.4 (3, 1432)	<0.001	3.02 (3, 1432)	0.029
Stress	0.002 (1, 1432)	0.964	41.9 (3, 1432)	<0.001	1.31 (3, 1432)	0.270
Motivation	15.0 (1, 1432)	<0.001	337 (3, 1432)	<0.001	2.43 (3, 1432)	0.064

TABLE IV. Gender differences in raw experience across activities and affective constructs. Abbreviations in the first column are affective construct (Aff), self-efficacy (SE), activation (Act), intrinsic motivation (Mot) and stress (Str).

Aff	Female			Male			$d$	$p$
	Mean	SD	$N$	Mean	SD	$N$		
<i>Physics</i>								
SE	1.57	0.82	82	2.23	0.76	148	0.77	(<0.001)
Act	1.99	0.58	82	2.13	0.70	148	0.22	0.09
Mot	1.25	0.56	82	1.61	0.69	148	0.53	(<0.001)
Str	1.48	1.12	82	1.30	0.97	148	-0.18	0.22
<i>STEM</i>								
SE	2.25	0.73	107	2.41	0.71	126	0.22	0.09
Act	1.99	0.63	107	1.97	0.69	126	-0.03	0.84
Mot	1.47	0.60	107	1.64	0.65	126	0.27	0.04
Str	0.92	0.92	107	0.95	0.87	126	0.04	0.78
<i>Non-STEM</i>								
SE	2.36	0.89	62	2.20	0.75	99	-0.20	0.24
Act	1.77	0.67	62	1.73	0.74	99	-0.06	0.71
Mot	1.53	0.91	62	1.68	0.70	99	0.19	0.27
Str	0.85	0.99	62	0.92	0.78	99	0.08	0.64
<i>Nonschool</i>								
SE	2.88	0.92	326	2.80	0.69	490	-0.10	0.21
Act	1.62	0.82	326	1.87	0.81	490	0.31	(<0.001)
Mot	2.76	0.77	326	2.81	0.76	490	0.07	0.30
Str	0.59	0.90	326	0.68	0.82	490	0.10	0.15

their experiences in physics were above their average overall experience,  $Z$  score = 0. Women's experiences in physics were amongst their worst self-efficacy experiences overall, with a rank of 21%. Men's mean self-efficacy experiences in physics ranked 14 points higher at 35%. Furthermore, with the exception of women's mean self-efficacy states in physics all other mean self-efficacy states in school activities ranked between 35% and 42%; a range half the size of the difference between men's and women's ranks in physics. Men were also two and a half times as likely to have above average self-efficacy states in the focal IE physics course—28% for men versus 11% for women. And the 17% difference in above average self-efficacy states between men and women in physics was much larger than the 10% range, 28%–38%, of self-efficacy states experienced above average in all school activities excluding women's experiences in the focal IE physics course.

Female students primarily experienced the IE physics course with low self-efficacy, whereas male students experiences tended toward high self-efficacy, Fig. 5. Approximately 1 in 4 of women's experiences were very low self-efficacy, whereby women experienced little to no control, success, or skill. Less than 10% of male students' experiences fell into this very low self-efficacy category. Female students had almost no experiences, 1%, of very high self-efficacy states whereas male students had 14% of their experiences be very high self-efficacy. These differences in

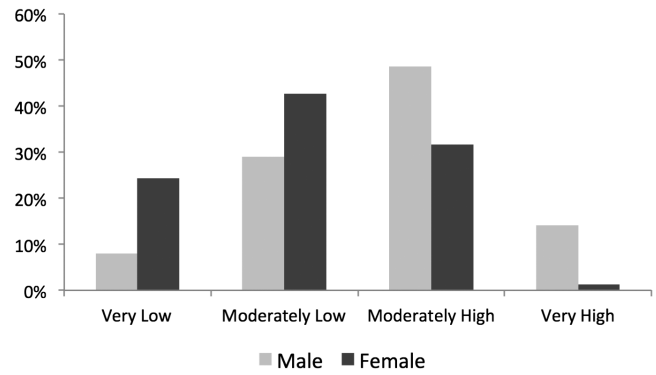


FIG. 5. Distribution of self-efficacy state experiences in IE physics by gender.

the distribution of experience provide further evidence that women experienced much lower levels of self-efficacy states in the focal IE physics course than their male peers did.

### C. Representativeness of trait participants

Male trait participants had higher mean grades ( $M = 2.69$ ,  $SD = 1.28$ ) than male nonparticipants ( $M = 2.10$ ,  $SD = 1.28$ ). Female trait participants had higher mean grades ( $M = 2.78$ ,  $SD = 1.26$ ) than female nonparticipants ( $M = 2.05$ ,  $SD = 1.16$ ). There were only small differences between males and females within the trait participant and nonparticipant groups. Univariate analysis of variance with a dependent variable of course grade and independent variables for trait participation and gender was statistically significant for participation  $F(1, 218) = 8.05$ ,  $p = 0.005$  but not for gender,  $p = 0.94$ , or the gender  $X$  participation interaction,  $p = 0.77$ . Similar to Kost *et al.* [3], those students who completed all trait measures and who make up the data set used to analyze gender differences overrepresent high achieving students and this trend was similar for male and female students.

### D. Gender differences in the focal IE physics course for trait measures

Male students started the course with slightly higher self-efficacy traits (3.47) than female students (3.29), Table V. Self-efficacy traits decreased for both male and female students with a very small shift for male students to a mean of 3.43, and a small shift for female students to a final mean of 3.13. These shifts were small, but the larger negative shift for women was consistent with the much worse self-efficacy states experienced by women in the IE physics course. The larger negative shift for self-efficacy traits for women resulted in the gender gap increasing a small amount from  $d = 0.34$  to  $d = 0.47$ .

Consistent with the gender differences for self-efficacy traits, male trait participants' mean scores were higher than females' on the pre- and post-measures for all other measures except for course grades. The female trait participants had slightly higher grades than the male



TABLE V. Gender differences in trait measures for the focal IE physics course and the courses studied by Kost *et al.* [3,7].

	Scale	Male			Female			<i>d</i> (95% CI)
		Mean	<i>N</i>	SD	Mean	<i>N</i>	SD	
<i>Focal IE Physics Course</i>								
Pre-FMCE	%	30.0	90	23.0	25.5	27	18.0	0.20 (−0.23, 0.63)
Post-FMCE	%	64.3	90	27.9	52.8	27	27.7	0.41* (−0.02, 0.84)
Pre-CLASS	%	64.1	90	15.2	57.9	27	18.5	0.39** (−0.05, 0.82)
Post-CLASS	%	60.8	90	17.5	52.8	27	19.4	0.44** (0.01, 0.87)
Pre-PSES	1–5	3.47	90	0.55	3.29	27	0.54	0.34 (−0.10, 0.77)
Post-PSES	1–5	3.43	90	0.69	3.13	27	0.46	0.47* (0.03, 0.90)
Course Grade	0–4	2.69	90	1.28	2.78	27	1.23	−0.07 (−0.50, 0.36)
<i>Other IE Physics Courses</i>								
Pre-FMCE	%	32.2	1566	21	22	533	16	0.51 (0.41, 0.61)
Post-FMCE	%	67.3	1566	27	56.8	533	29	0.38 (0.28, 0.48)
Pre-CLASS	%	65.7	1380	16	63.6	522	15	0.13 (0.03, 0.23)
Post-CLASS	%	56	...	...	52.3	...	...	...
Pre-PSES	1–5	...	...	...	...	...	...	0.29
Post-PSES	1–5	...	...	...	...	...	...	0.36
Course Grade	0–4	2.53	2715	0.99	2.41	848	0.92	0.11 (0.03, 0.19)

\* $p < 0.05$ , \*\* $p < 0.01$ .

participants, 2.78 versus 2.68. Gender differences on all measures except course grades were small to medium in size, 0.20 to 0.47, favored male students, and increased from pre- to postmeasurement.

We used a 2X7 omnibus MANOVA to identify if there was a main effect of gender on trait measures for the trait participants. A secondary purpose was to identify if there were also effects for participating in the ESM and for ESM X gender interaction that were discussed earlier. Independent variables were student gender and participation in the ESM data. Dependent variables were course grade and the pre- and postmeasures for the FMCE, CLASS, and PSES. The MANOVA showed a statistically significant difference for gender  $F(7, 107) = 2.85$ ,  $p = 0.009$ . Statistically significant gender differences identified by the subsequent factorial analysis of variance are indicated in Table V.

The results for trait measures for both overall scores and gender differences in scores were consistent with the results reported for other IE physics courses [3,7]. Means for the measures were mostly very similar,  $d < 0.2$ , and there was no consistent pattern of one course having higher means than the other course. For women the differences greater than 0.2 standard deviations were small for the post FMCE,  $d = 0.22$ , and moderate for both the pre-CLASS,  $d = 0.38$ , and course grades,  $d = 0.40$ . With the women in the focal course having higher grades and post FMCE scores but lower CLASS scores. For male students the only noteworthy difference was the small effect on the post CLASS,  $d = 0.27$ . All other differences were very small. These results indicate mostly small and inconsistent differences

between the two courses and indicate that the students in the courses started and ended instruction similarly.

The gender differences and shifts in gender differences were also similar in the two courses. All of the differences favored male students and increased from pre- to post-instruction. While the gender differences on the PSES were very similar between the focal and other IE courses, those on the CLASS and FMCE had some variability. However, this variability can be explained by relatively small differences, on the order of one question, between the means for men and women in the two courses. Subsequently, we concluded that these courses had similar populations of students and that shifts in students' traits from pre- to postinstruction were similar.

## VI. DISCUSSION

While learning physics, women did not experience high self-efficacy states, as men sometimes did. Instead women frequently experienced low or very low self-efficacy states, and, correspondingly, their self-efficacy traits were significantly reduced from pre- to postcourse. Men, by contrast, had very small negative shifts in their self-efficacy traits, consistent with the higher levels of self-efficacy states that they experienced. Furthermore, there was no other activity in which either men or women had such low self-efficacy states as women experienced in the focal IE physics course. Supporting the validity of our self-efficacy measures is the finding that the gender differences we observed were consistent with gender differences on the complementary measures, both state and trait. This is especially true for

state measures, which showed that women experienced the focal IE physics course with less activation, more extrinsic motivation, and greater stress than men. For traits, similar to self-efficacy, gender differences increased for conceptual knowledge and attitudes about learning physics. Thus, our overall conclusion is that the larger negative shift in women's self-efficacy traits in the physics course was caused by the experience of instruction in which women's self-efficacy states were much worse than men's.

Earlier, we brought up three alternative explanations for the larger negative shift in women's self-efficacy traits: that they were the result of a broad trend across many college courses, that they were a result of differences in experience in marginal activities in physics learning, or that there was no difference in experience only a difference in retrospection. Our findings demonstrate all three alternative explanations had little to do with the disparate effects on women's self-efficacy. Most importantly, there was no indication that the larger negative shift in women's self-efficacy traits was a part of a larger trend. The large gender differences in self-efficacy states only occurred in the focal IE physics course and did not occur in other STEM courses. Second, while it is possible that women experience marginal activities in physics learning differently than men, the large differences that we measured for much more common activities make it unlikely that marginal experiences play a more important role than the experience of learning physics that we measured. Last, the large gender differences in the experience of learning physics ruled out the possibility that the larger negative shifts were due only to differences in retrospection.

Based on the similarity of the focal IE physics course to other IE courses and the consistency between the large gender differences in self-efficacy states and concurrent larger negative shift in women's self-efficacy traits in the focal course, we think it is probable that similar gender differences in self-efficacy states exist in other physics courses using either IE or traditional instruction. This is supported by most investigations, revealing that IE and traditional lecture physics courses had larger negative impacts on female students' self-efficacy traits [7–10]. To be sure, there may have been something idiosyncratic to the focal course that was depressing women's self efficacy in our study, something that would not be a regular feature of other IE physics courses. An important example of course idiosyncrasies are differences between instructors, which can cause large differences in how students experience otherwise similar courses [50]. Given that the present study focused on a single IE physics course we cannot firmly rule out the possibility of idiosyncrasies in that course uniquely affecting women's experiences. However, if idiosyncrasies uniquely affected women's self-efficacy states then the self-efficacy trait outcomes for the focal course should have been more severe than in other IE physics courses. In fact, the

self-efficacy trait outcomes for the focal IE physics course featured in this study were similar to other IE physics courses, suggesting that this course was representative of IE physics instruction in general. Furthermore, the focal course had similar conceptual learning outcomes to other courses implementing IE pedagogies [2,3,28] and had similar gender differences to other IE physics courses on all four trait measures including conceptual knowledge, self-efficacy traits, attitudes about learning physics, and course grades [3,7].

Although our favored explanation is that the IE physics instruction negatively impacted women's self-efficacy, two alternative explanations bear discussing. Both attribute the cause of the effects to the female students rather than to the learning environment. First, it may be that the gender differences in state experiences were not a result of gender *per se*, but rather a result of the trait factors that varied with gender, namely, female students' lower conceptual knowledge, less expertlike attitudes, and lower self-efficacy traits at the beginning of the course. Nissen [45] checked on this possibility using regression analysis to determine whether gender explained significant variance in mean self-efficacy states when controlling for trait effects. The results were that gender was the largest predictor of an individual's average self-efficacy state experiences while controlling for traits. Thus, there was something about being a woman in physics, over and above the measured physics traits, that made the experience of IE physics harmful to women's self-efficacy. A second alternative explanation, particularly for the large gender difference in self-efficacy states, can be attributed to the recruitment process. It is possible that recruiting students in their IE physics class biased the sample by attracting female volunteers who were particularly interested in having their negative experiences in physics be understood. The recruitment process did not indicate that the study was about students' emotions or feelings to limit this possibility. Nevertheless, students could easily have inferred that our study of "experiences" would include affect, so the possibility of sample bias cannot be ruled out. The results reported for self-efficacy traits, however, provide some evidence that, if there was sample bias, it was in a conservative direction. Namely, female ESM participants had smaller negative shifts pre- to postcourse than female non-ESM participants and women's low self-efficacy states were evenly distributed across a sample of 1 in 4 women in the course who overrepresented higher-achieving women. Thus, even if the sample were biased, these results apply to a significant subpopulation of women taking physics.

What is it about IE physics instruction that is harmful to women's self-efficacy? STEM courses naturally have many features in common with physics in terms of physical environment, course structure, assignment of grades, and so on. It is the nature of instruction and subject matter that varies most from course to course. One possibility is that

there is something about the content in physics that somehow makes women feel less efficacious than does the content of other introductory courses such as mathematics, chemistry, and engineering principles. Along these lines Taasobshirazi and Carr [51] suggested that women may be disadvantaged in physics because of its emphasis on spatial thinking, which interacts with gender differences in spatial ability. The results of the present study are only partially consistent with this interpretation in that men did exhibit higher scores on conceptual knowledge than women. However, contrary to this interpretation, women ended the course similar to men with respect to grades. Thus, it is not obvious that men were leveraging their presumably higher average spatial ability very well. Meanwhile, as reported earlier in this article Modeling Instruction physics courses tentatively do not negatively impact students' self-efficacy traits or provide for differential shifts between men and women. Therefore, we think that pedagogy likely plays a larger role than subject matter in the observed effects within IE Physics.

## VII. CONCLUSION

Here we have used the ESM to situate the experiences of interest, self-efficacy in physics, within the breadth of experience while minimizing the effects of retrospection. This demonstrated how the ESM can be scaled to capture a large collection of experiences across a broad range of activities for a large number of participants. These features complement and bridge the fine-grained detail that can be achieved with case studies using interviews or video analysis and the large-scale data that surveys provide. The present study did not allow us to conjecture about which aspects of physics pedagogy substantially impacted student's experiences, however, different designs leveraging the strengths of the ESM in combination with other

methods can achieve this goal. For instance, a useful form of research to identify the possible causal relationships and mechanisms between specific aspects of instruction and student experience would be case studies detailing the experiences of ESM participants whose self-efficacy states fell in either of the extremes. A second approach would be to use a similar design to the present study, but with a larger sample of experience, to increase the resolution of the ESM. This would allow investigating self-efficacy states within specific aspects of instruction, such as answering ConcepTests in lecture, and linking these experiences to students' shifts in physics self-efficacy traits. Both of these approaches would benefit from collecting data in multiple courses and across different pedagogies.

The poor experiences, poor outcomes, and underrepresentation of women in physics warrants future research to inform addressing and resolving these issues. Many students leave STEM and physics because of their poor experiences and despite being fully capable of succeeding in the material [18]. By physics instruction undermining women's self-efficacy traits, physics instruction is also likely undermining women's performance, persistence, and selection of physics as a major. Here we have shown that self-efficacy is important to understanding the underrepresentation of women in physics. Developing physics instruction that supports positive self-efficacy states is a starting point for instruction that supports all students in meeting both the affective and the cognitive demands of learning physics, especially the development of self-efficacy traits. Such instruction is necessary for physics to inclusively support diverse populations of students [13,52]. Otherwise, physics is likely to continue to lag behind other STEM disciplines in diversity, threatening its survival as a major subject of study as it becomes an anachronism in an ever more diverse world.

- 
- [1] National Science Foundation, Women, minorities, and persons with disabilities in science and engineering (2012), <http://www.nsf.gov/statistics/2015/nsf15311/>.
  - [2] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020121 (2013).
  - [3] L. Kost, S. Pollock, and N. Finkelstein, Characterizing the gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009).
  - [4] A. Bandura, *Self-Efficacy: The Exercise of Control* (Longman, New York, 1997).
  - [5] K. Multon, S. Brown, and R. Lent, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, *J. Counsel. Psychol.*, **38**, 30 (1991).
  - [6] V. Sawtelle, E. Brewé, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *J. Res. Sci. Teach.* **49**, 1096 (2012).
  - [7] L. Elizabeth Kost-Smith, University of Colorado, Characterizing, Modeling, and Addressing Gender Disparities in Introductory College Physics, Ph.D. thesis 2011.
  - [8] V. Sawtelle, E. Brewé, and L. H. Kramer, Positive Impacts of Modeling Instruction on Self-Efficacy, in *PERC Proceedings* (AIP, Melville, NY, 2010), pp. 289–292.
  - [9] A. Cavallo, M. Rozman, and W. Potter, 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).



- [10] C. Lindström and M. Sharma, Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?, *Int. J. Innovation Sci. Math. Educ.* **19**, 1 (2011).
- [11] A. Bandura, Guide for constructing self-efficacy scales, in *Self-Efficacy Beliefs of Adolescents*, edited by F. Pajares and T. Urdan (Information Age Publishing, Charlotte, NC, 2006), pp. 307–337.
- [12] H. S. 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).
- [13] E. Redish, J. Saul, and R. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- [14] E. Brewé, L. Kramer, and G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS, *Phys. Rev. ST Phys. Educ. Res.* **5**, 013102 (2009).
- [15] L. Kost-Smith, S. Pollock, and N. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a smog of bias, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020112 (2010).
- [16] V. Sawtelle, E. Brewé, R. Michelle Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020111 (2012).
- [17] R. M. Hall and B. R. Sandler, *The Classroom Climate: A Chilly One for Women* (PSEW, AAC, Washington DC, 1982).
- [18] E. Seymour and N. Hewitt, *Talking about Leaving: Why Undergraduates Leave the Sciences* (Westview, Boulder CO, 1997).
- [19] T. Mujtaba and M. J. Reiss, Inequality in experiences of physics education: Secondary school girls' and boys' perceptions of their physics education and intentions to continue with physics after the age of 16, *Int. J. Sci. Educ.* **35**, 1824 (2013).
- [20] L. Kost, S. Pollock, and N. Finkelstein, Unpacking gender differences in students' perceived experiences in introductory physics, in *PERC Proceedings* (2009), pp. 1–3.
- [21] J. S. Hyde, E. Fennema, M. Ryan, L. A. Frost, and C. Hopp, Gender comparisons of mathematics attitudes and affect: A meta-analysis, *Psychol. Women Q.* **14**, 299 (1990).
- [22] T. Goetz, M. Bieg, O. Lüdtke, R. Pekrun, and N. C. Hall, Do girls really experience more anxiety in mathematics?, *Psychol. Sci.* **24**, 2079 (2013).
- [23] M. Bieg, T. Goetz, and A. a Lipnevich, What students think they feel differs from what they really feel—academic self-concept moderates the discrepancy between students' trait and state emotional self-reports, *PLoS One* **9**, e92563 (2014).
- [24] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, Englewood Cliffs, NJ, 1986).
- [25] J. Jackson, P. Hill, and B. Roberts, Misconceptions of traits continue to persist: A response to Bandura, *Journal of management* **38**, 745 (2012).
- [26] D. Anthony Luzzo, P. Hasper, K. Albert, M. Bibby, and E. Martinelli, Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, *J. Counsel. Psychol.* **46**, 233 (1999).
- [27] T. Williams and K. Williams, Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations, *J. Educ. Psychol.* **102**, 453 (2010).
- [28] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [29] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [30] C. Turpen and N. Finkelstein, Not all interactive engagement is the same: Variations in physics professors implementation of Peer Instruction, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020101 (2009).
- [31] E. Mazur, *Peer Instruction: A Users Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [32] L. McDermott and P. Schaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2002).
- [33] V. Otero, S. Pollock, and N. Finkelstein, A physics departments role in preparing physics teachers: The Colorado learning assistant model, *Am. J. Phys.* **78**, 1218 (2010).
- [34] W. Adams, K. Perkins, N. Podolefsky, M. Dubson, N. Finkelstein, and C. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado learning attitudes about science survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
- [35] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- [36] R. Thornton, D. Kuhl, K. Cummings, and J. Marx, Comparing the force and motion conceptual evaluation and the force concept inventory, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010105 (2009).
- [37] M. Wittmann, FMCE Grading Template (2009), <http://perlnet.umephy.maine.edu/materials/>.
- [38] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, *Experience Sampling Method: Measuring the Quality of Everyday Life* (SAGE Publications, Thousand Oaks, CA, 2007).
- [39] R. E. Thayer, *The origion of everyday moods: Managing energy, tension and stress* (Oxford University Press, New York, 1996).
- [40] R. E. Thayer, Measurement of activation through self-report, *Psychological reports* **20**, 663 (1967).
- [41] E. Deci, R. Koestner, and R. Ryan, A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, *Psychol. Bull.* **125**, 627 (1999).
- [42] M. Csikszentmihalyi, *Beyond Boredom and Anxiety* (Jossey-Bass Publishers, San Francisco, CA, 1975).
- [43] R. Lazarus and S. Folkman, *Stress, Appraisal and Coping* (Springer, New York, 1984).
- [44] J. A. Schmidt and L. Shumow, Change in Self-Efficacy in High School Science Classrooms, in *Self-Efficacy in*

- School and Community Settings*, edited by S. Britner (Nova Science Publishers, New York, 2012), pp. 53–73.
- [45] J. Nissen, Self-efficacy state experiences and trait development: A gender study in an interactive engagement physics course, Ph.D. thesis, University of Maine (2016).
- [46] J. Michael Nissen and J. T. Shemwell, Measuring affective experience in the midst of STEM learning, *Proceedings of the Eleventh International Conference for the Learning Sciences (2014)* (International Society of the Learning Sciences, Boulder, 2014), Vol. 1, pp. 142–149.
- [47] R. Ryan and E. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, *Am. Psychol.* **55**, 68 (2000).
- [48] I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer, Impact of equity models and statistical measures on interpretations of educational reform, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020103 (2012).
- [49] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, 1977).
- [50] W. L. Sanders and S. P. Horn, The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment, *J. Personnel Evaluation Educ.* **8**, 299 (1994).
- [51] G. Taasobshirazi and M. Carr, Gender differences in science: An expertise perspective, *Educ. Psychol. Rev.* **20**, 149 (2008).
- [52] Z. Hazari, R. H. Tai, and P. M. Sadler, Introductory university physics performance: The influence of high school physics preparation, *Sci. Educ.* **91**, 847 (2007).