Impact of traditional or evidence-based active-engagement instruction on introductory female and male students' attitudes and approaches to physics problem solving

Melanie Good,¹ Alexandru Maries,² and Chandralekha Singh¹

¹Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA ²Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA

(Received 12 May 2018; published 20 September 2019)

Over the course of instruction, not only does most introductory physics students' content knowledge evolve but their attitudes and approaches to problem solving are also likely to evolve. This change may depend on many factors including the curricula and pedagogies used, the degree to which instruction actively engages students in the learning process, as well as the gender of the students. While changes in epistemology and beliefs about physics have been examined in the literature, how students' attitudes and approaches to problem solving change from the beginning to the end of instruction in introductory physics and how method of instruction or gender of the student impact them remain largely unexplored. To examine the potential changes in attitudes and approaches to problem solving over a semester, we administered a previously validated attitudes and approaches to problem solving (AAPS) survey both at the beginning (pre) and at the end of instruction (post) in eight large enrollment calculus-based introductory physics classes at a large research university in the United States. At both points in time (beginning and end of a semester of instruction), each class was also given surveys measuring students' conceptual understanding: the Force Concept Inventory (FCI) or the Conceptual Survey of Electricity and Magnetism (CSEM), depending upon whether it was the first or second semester course. In addition, final exam scores, gender of students, and descriptions of the instructional methods used for each class were collected. The AAPS survey was used to measure students' attitudes and approaches to problem solving, and the conceptual surveys and exam scores were used to measure the degree to which each course helped students learn physics concepts. We examined students' performance on the AAPS survey, FCI or CSEM, and final exams, and compared the results for different instructional methods, and gender of students. Moreover, we examined whether or not there were correlations between the expertlike response on the AAPS survey and the performance on FCI or CSEM or final exams. We found that all classes exhibited a decline in score on the AAPS survey suggesting worse attitudes related to problem solving after instruction. Furthermore, controlling for the initial scores, classes which involved significant use of evidence-based active engagement methods exhibited statistically significantly better scores on the AAPS survey at the end of the course compared to classes which were taught primarily using a traditional lecture-based approach. Equally importantly, unlike broader epistemological surveys, female students were found to exhibit less of a decline in AAPS scores than did their male counterparts in all classes and the AAPS scores were always higher for female students at the end of the course. Future research should contemplate how this novel finding may be effectively exploited to develop and implement curricula and pedagogies to reduce the gender gap in performance often observed in introductory physics.

DOI: 10.1103/PhysRevPhysEducRes.15.020129

I. INTRODUCTION

Instructional goals of many introductory physics courses include helping students develop expertlike problem solving skills while learning physics concepts. Therefore, many physics education researchers have been involved in investigating expert-novice differences in physics problem solving, how students develop their problem solving skills, and instructional strategies to make students perform more expertlike tasks, see, e.g., Refs. [1–37]. These instructional goals may be facilitated by helping students develop attitudes and approaches to problem solving that reflect the way an expert might think about and solve physics problems.

Reference [1] is a synthesis of discipline-based education research in physics and Ref. [2] is an overview of physics education research on problem solving. In particular, Chi *et al.* [3] conducted research showing differences in

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

the expert-novice problems solving via problem categorization and Reif *et al.* [4–6] investigated the effect of knowledge organization on task performance and prescribed effective problem solving processes.

Mestre and colleagues [7–10] have conducted many studies focusing on various facets of expert-novice differences, e.g., the relation between problem categorization and problem solving among experts and novices and how to make beginning physics students perform more expertlike tasks, e.g., by using qualitative problem solving strategies to highlight the role of conceptual knowledge in solving problems and constraining students to perform expertlike problem solving by schema acquisition. Mestre and collaborators [11-14] have also conducted research on transfer of learning, including research on conceptual understanding and transfer of learning via problem posing, the role of knowledge coordination in the transfer process, whether transfer is ubiquitous or rare, and developing a framework for transfer of learning as sense making. Harper [15] and Heller et al. [16–18] have investigated physics students' problem solving behaviors, advantages of cooperative group problem solving in physics and faculty beliefs and values about the teaching and learning of problem solving. Meltzer [19] investigated the relation between representational mode and physics problem solving and Mashood et al. [20] conducted large-scale studies on the transferability of general problem solving skills. Gladding et al. [21] investigated student learning using mastery style vs immediate feedback online activities and Ding et al. [22] have used conceptual scaffolding to foster effective problem solving in introductory physics.

Our group [23–30] has been involved in researching the role of intuition in physics problem solving, evaluating expertise of introductory physics students using isomorphic problems, investigating challenges in using analogies to help students develop problem solving skills, the effect of scaffolding on helping students solve quantitative problems involving strong alternative conceptions, and using isomorphic problem pairs to help students learn to transfer their learning from one context to another. We have also been involved [31–37] in investigating the role of asking students to diagnose their mistakes by providing incentives to learn from mistakes, using categorization of problems as an instructional tool, challenges in designing appropriate scaffolding to improve students' representational consistency, and developing interactive web-based tutorials to develop expertise in introductory physics students.

Here we discuss an investigation in which a validated survey was used to probe changes in calculus-based introductory physics students' attitudes and approaches to problem solving from the beginning to the end of the course at a large research university. Since evidence-based active engagement pedagogies can impact students' attitudes and approaches to problem solving, we compared the changes in students' attitudes and approaches to problem solving in courses employing these types of pedagogies with those employing traditional lecture-based instruction. Moreover, prior research suggests that female students in an introductory physics course on average have worse scores than male students on broader physics epistemological surveys after instruction [38]. We investigated whether gender trends are similar or different for the survey focused only on attitudes and approaches to problem solving when data in different types of courses are analyzed separately for male and female students. In addition, since investigation of the correlation between male and female students' attitudes and approaches to problem solving and their actual physics performance can be a valuable measure of students' expertise and can be helpful in developing pedagogical strategies to improve student learning, we investigated it.

A. Expert vs novice problem solving

Physics experts, e.g., physics faculty members, organize their physics knowledge hierarchically so that the underlying concepts are connected in a meaningful and structured way in schema and they generally exhibit positive attitudes and employ effective approaches to scientific problem solving [1–37]. Experts' knowledge is structured in well-organized schema and their positive attitudes and effective approaches to problem solving can facilitate desired problem solution [3-13,39-42]. By contrast, novices, e.g., many introductory students, do not possess a wellorganized knowledge structure and they view physics as a collection of disconnected facts and equations, and often have less expertlike attitudes and approaches to problem solving. When asked to group together physics problems based upon their solutions [3], physics experts often group them based upon the deep features, e.g., the physics principles involved while novices often group them based upon the surface features (e.g., they often group all inclined-plane problems together in one group and all spring problems in another group even if different inclined-plane problems are solved using different physics principles).

An effective approach to physics problem solving involves starting with a conceptual analysis of the problem by contemplating on the information provided, reflecting upon the goals, transforming the initial problem representation to another representation that makes the analysis easier, e.g., by drawing a diagram and predicting some features of the solution. After this initial stage of problem solving, an expert generally plans the solution by dividing the problem into subproblems and considering what physics principles and concepts may be applicable for each subproblem. Drawing analogies with familiar situations and considering limiting cases are expertlike problem solving heuristics at the initial qualitative analysis and planning stages of problem solving [23]. Generally, only after these stages of problem solving does an expert implement the plan and then reflects upon the problem solving process. The reflection phase of the problem solving involves performing a reasonability check to ensure that the answer makes sense (e.g., whether the units are correct based upon the physical quantity calculated or the speed of a car is not more than the speed of sound) and also doing metacognition about the deep features of the problem (e.g., why a particular principle of physics was useful in this context and how one would recognize that the same principle should be used in a future physics problem that may have a different context). If the initial solution to a problem is not reasonable, the expert problem solving approach involves an iterative process and reflection on where the error may have cropped up [23]. Physics experts are likely to learn from their own mistakes and use problem solving as an opportunity to repair and extend their knowledge structure and not make the same mistake again. Unfortunately, without explicit guidance and incentive, many introductory physics students approach physics problem solving in a haphazard manner. Many use a "plug-and-chug" approach and look for physics formulas even before conceptually analyzing the problem and planning a solution, and after solving the problem, they often skip the reasonability check. When correct solutions are provided, they do not necessarily learn from their mistakes [31,32].

However, in order to use problem solving to develop a robust knowledge structure and become a physics expert, students must learn to combine conceptual and quantitative aspects of problem solving and regard physics principles encapsulated in compact mathematical forms as relations between physical quantities and reflect upon how those fundamental principles are applicable in the situation posed (instead of merely using them as a plug-and-chug tool to obtain a solution without regard to the underlying concepts). In particular, contrary to the plug-and-chug approach used by many novice problem solvers, conceptual and quantitative aspects of problem solving should be intertwined in order to solve a physics problem effectively as well as to learn from problem solving and using it to organize, extend, and repair one's knowledge structure. For example, one has to analyze or apply physics principles appropriately to solve a problem meaningfully at the conceptual analysis and planning stages of problem solving regardless of whether it is a problem involving a single principle of physics or a problem requiring synthesis of several physics principles [2,22,23]. Similarly, problem solvers who use an expertlike approach to problem solving also make an explicit connection between the conceptual and quantitative aspects of the problem at the reflection stage of problem solving and use this connection to extend and repair their knowledge structure (if there was an error in their initial solution or if they struggled to solve the problem). Employing these effective problem solving heuristics consistently in diverse situations can increasingly make the problem solvers adaptive experts in that they may become adept at applying

different physics principles and concepts to solve diverse problems in unfamiliar situations. Prior studies suggest that providing explicit guidance and/or incentives to use effective problem solving strategies and learn from mistakes can help students develop expertise in problem solving [31,32].

B. Evaluating growth among introductory physics students along various dimensions using surveys before (pre) and after (post) instruction

One can measure changes along various dimensions from the beginning to the end of a course as a result of instruction, e.g., growth in content knowledge, and/or approaches to problem solving. Students' epistemological beliefs represents one dimension that can impact learning in a particular discipline [38,43–46]. Several surveys have been developed to evaluate both students' attitudes about physics and physics learning [38,44–48] as well as their conceptual understanding of physics. Attitudinal surveys that focus on students' epistemological beliefs about physics include the Maryland Physics Expectation Survey (MPEX) and the Colorado Learning Attitudes about Science Survey (CLASS) [38,46]. When these attitudinal surveys have been administered at both the beginning and end of a semester of instruction, the scores typically show declines in students' attitudes towards physics and physics learning over the course of instruction compared to what would be considered an expertlike response [47]. This has been found to be true for both male and female students, even when students could correctly identify how their instructors would have answered the survey questions [38].

The Attitudes toward Problem Solving Survey (APSS) was developed with inspiration from the MPEX survey with a focus only on attitudes about problem solving [44,48]. The APSS reveals a similar decline in student attitudes from the beginning to the end of the semester after instruction to that found via the MPEX and CLASS surveys, especially in large enrollment traditionally taught classes [44]. Here we focus on the attitudes and approaches to problem solving survey [49–52] which is a modified version the APSS survey.

In addition to attitudinal surveys, several conceptual surveys have been developed and validated to assess students' conceptual understanding. Conceptual surveys include, among others, the Force Concept Inventory (FCI) [53–57], which is usually given to first semester introductory physics students learning mechanics, and the Conceptual Survey of Electricity and Magnetism (CSEM) [58], which is typically administered to second semester introductory physics students learning electricity and magnetism.

The Force Concept Inventory (FCI) was developed to assess students' conceptual understanding of introductory Newtonian mechanics [53–57]. The FCI has been administered to introductory physics students in various universities in the United States and elsewhere for decades.

The Conceptual Survey of Electricity and Magnetism (CSEM) was developed to assess students' conceptual understanding of introductory electricity and magnetism topics broadly [58]. The CSEM is a particularly difficult survey for introductory students, with typical scores being lower than 50% [58] after instruction in relevant concepts.

C. Evidence-based active engagement (EBAE) methods

Many introductory physics classes in the U.S. are taught primarily in a lecture-based manner. Instruction in which regular class time is primarily focused on the instructor lecturing and the students taking notes is often referred to as "traditional." However, some physics faculty members implement physics education research-based instructional strategies (RBIS) in their teaching [59]. There is growing awareness of RBIS among physics faculty, especially among those full-time faculty who attend teaching-related workshops and/or read teaching-related journals [60,61]. We refer to RBIS as "evidence-based active engagement" (EBAE) methods. The EBAE methods may vary significantly, but they share a common goal: facilitate an environment in which students take a more active role in their learning using physics education research-based approaches than is afforded by traditional approaches [62]. These EBAE courses not only have the potential to impact introductory physics students' content knowledge and skills but also their attitudes and approaches to problem solving.

Peer Instruction is an example of an EBAE method in which in-class clicker questions along with student discussion are interspersed throughout instruction [63]. This method has been shown to promote both conceptual understanding and better problem solving skills, and may also positively impact students' attitudes about physics and physics learning [63-65]. Another EBAE method that has been associated with conceptual learning gains [66] involves the use of interactive lecture demonstrations (ILDs) in which students are asked to predict what will happen before a demonstration is shown during class. After the demonstration, the instructor leads a class discussion guiding students to build a coherent knowledge structure of the concepts involved. Collaborative group work involving the use of context-rich problems is yet another example of an active engagement method in the physics classroom [16,17]. Use of context-rich problems in collaborative group work has been associated with the development of expertlike problem solving approaches [16,17]. Moreover, "inverting" or "flipping" instruction is becoming increasingly popular and provides an opportunity to move some of the content which is normally covered in lectures to assigned videos and/or tutorials outside of the class [67–69]. This allows more class time to be used for EBAE activities (with students often working in small groups) that are designed to reinforce the content presented outside of class using videos, textbooks, and other resources [67].

D. Gender differences in introductory physics performance

Prior research suggests that there is often an average performance gap based upon the gender of the students who take the FCI, with male students scoring higher, on average, than female students [68,70–72]. The discrepancy between male and female students' performance is commonly referred to as the "gender gap" [68,70-72] and has been found even after controlling for factors such as different prior preparation or coursework of male and female students [71,72]. Like the FCI, male and female students often perform differently on the CSEM, but the gender gap is typically not as large for the CSEM as it is for the FCI [57]. Moreover, findings are mixed as to the degree to which the use of the EBAE methods impacts the gender gap on conceptual surveys such as the FCI and CSEM. Some prior research has found that using carefully designed evidence-based pedagogies can reduce the gender gap on standardized conceptual assessments [57], but other studies suggest that a gender gap remains even when EBAE methods are used [68,73-75].

The origin of the gender gap on the FCI both at the beginning and end of a physics course has been a subject of debate, raising questions about whether the test itself may be gender biased [70]. Some of the origins of the gender gap can be attributed to societal gender stereotypes [76,77] that begin from an early age and cause stereotype threats for female students. For example, research suggests that even six year old boys and girls have gendered views about intelligence in which they view boys as smarter [76]. Such stereotypes can impact female students' self-efficacy [78], their beliefs about their ability to perform well in disciplines such as physics in which they are underrepresented and which have been associated with "brilliance."

E. Research focus and framework for our investigation

Our research focused on examining the changes in the student response to the attitudes and approaches to problem solving (AAPS) survey from the beginning to the end of a semester in introductory calculus-based physics courses at a large state-related research university in the Unites States. Furthermore, we investigated these changes for different instructional approaches (EBAE vs traditional lecture-based) and for female and male students separately. Moreover, we investigated student performance on standardized conceptual surveys or final exam and the correlation between student performance on conceptual surveys or final exam and the AAPS survey.

It is important to investigate how the attitudes and approaches to problem solving using the AAPS survey are impacted by whether primarily traditional lecture or EBAE methods are employed. The overarching framework that inspired the comparison of the instructional methods in this study is that the EBAE methods focused on a cognitive approach to instructional design and building on students' prior knowledge to facilitate learning and development of a robust knowledge structure. In these EBAE methods, there was an effort to align learning goals and objectives, instructional design, and assessment of learning with each other and there was a focus on evaluating whether the pedagogical approaches employed have been successful in meeting the goals and enhancing student learning. The instructors using the EBAE approaches were employing the cognitive apprenticeship model [79], which focuses on "modeling," "coaching and scaffolding," and "weaning." In particular, providing opportunities to coach students while they engage with problem solving and scaffold their learning was a central aspect of the EBAE methods used in our investigation. As noted in the introduction, problem solving with functional understanding (as opposed to a plug-and-chug approach) requires conceptual understanding of the relevant physics concepts and principles. Also, a solid conceptual understanding can reduce the cognitive load [80] while solving problems. Moreover, since quizzes and exams that were common in all of the courses primarily involved problem solving in which students had to apply and synthesize physics principles and concepts to solve complex problems and less than 10% of assessments were conceptual, even the EBAE courses focused primarily on problem solving even though conceptual clicker questions were used in such courses to varying degrees to strengthen conceptual understanding. Since no prior research has focused on the differences between student attitudes and approaches to problem solving in these types of EBAE courses compared to those with teaching focused primarily using traditional lecture, we investigated it.

Moreover, prior research suggests that the way in which physics is taught has been found to be connected to students' beliefs about physics [81]. Therefore, it is important to explore whether attitudes and approaches to problem solving in physics are likewise connected to the method of instruction. While other attitudinal surveys have investigated broader questions regarding epistemology and beliefs about physics [38,46,47], and how these beliefs may be related to methods of instruction [72], no study has focused on the aspect of student perspectives related specifically to attitudes and approaches to problem solving, and whether or not the method of instruction is related to these attitudes and approaches. Our hypothesis is that EBAE instructional strategies that focus on helping students develop effective problem solving skills may encourage better attitudes and approaches to problem solving. We therefore investigated the impact of instruction on the AAPS scores at the end of the semester.

Furthermore, the broader epistemological CLASS survey shows that female students overall score worse than male students, but it would be beneficial to investigate gender differences on the AAPS survey, which focuses

specifically on the attitudes and approaches to problem solving [38]. In addition, prior research suggests that female students in introductory physics courses for science and engineering majors, on average, have lower selfefficacy and intelligence mindset than male students even after controlling for performance [82-90] and trends are often similar for racial and ethnic minority students [91]. Also, activation of a stereotype, i.e., stereotype threat, about a particular group in a test-taking situation can alter the performance of that group in a way consistent with the stereotype. Some researchers have argued [77] that female students, when working on a physics test, undergo an implicit stereotype threat due to the prevalent societal stereotypes. However, no prior research has focused on differences in attitudes and approaches to problem solving of female and male students in introductory physics courses, which may prove to be useful in improving the learning environment for all students in introductory physics.

Furthermore, since investigation of correlation between male and female students' AAPS survey scores and their actual performance on the conceptual assessment or final exam in EBAE and traditionally taught courses can be a valuable measure of students developing physics expertise and can even be helpful in developing pedagogical strategies to improve student learning, we investigated it. As discussed in the introduction, since conceptual understanding is important for solving quantitative problems with functional understanding (as opposed to when using a plug and chug approach), correlation between AAPS survey scores and both conceptual standardized test and final exam performances were examined.

F. Research questions

Our research questions are as follows:

RQ1. How do the presurvey (i.e., before instruction in relevant concepts) and postsurvey (i.e., after instruction in relevant concepts) scores on the AAPS survey compare? And how do those measures differ by course (i.e., Physics 1 and Physics 2)?

RQ2. How do the AAPS survey scores in physics 1 compare for EBAE vs traditional instruction without separating students by gender and how do they compare when male and female students are considered separately?

RQ3. How correlated are students' AAPS survey scores in a given class with their conceptual survey (FCI or CSEM) or final exam performance?

II. METHODOLOGY

A. Validation of the attitude and approaches to problem solving survey

We employed a modified version of the APSS, the AAPS survey, which was developed and validated to include questions regarding the approaches students take when solving physics problems [49–51]. The AAPS survey is

unique in its focus compared with other broader attitudinal survey because, like the APSS, it hones in specifically on problem solving, but unlike the APSS, the AAPS survey also probes approaches to problem solving in addition to attitudes towards problem solving. The added dimension of approaches to problem solving is important for assessing growth in problem solving expertise. To investigate the evolution of expertlike response, the AAPS survey was validated based upon introductory student, graduate student, and faculty responses [49]. Since physics experts, e.g., physics faculty members, organize their physics knowledge hierarchically with underlying concepts connected in a meaningful and structured way, their attitudes and approaches towards problem solving, as reflected by their response to the AAPS survey, establish the criteria for an expertlike response. By contrast, many introductory students have less expertlike attitudes and approaches to physics problem solving [49]. After initial validation [49,50], the AAPS survey has been used to investigate students' attitudes and approaches to problem solving in high school and university physics classes in Turkey and South Africa [51,52]. However, the AAPS survey has not been used to investigate changes in introductory physics students' attitudes towards problem solving before and after instruction in an introductory physics course nor has it been used to investigate gender differences or impact of different methods of instruction.

We note that although the AAPS survey is a validated survey, in the past, it has only been administered toward the end of the course to introductory college physics students and high school students [49-51]. Therefore, we conducted audio-recorded interviews with seven students before they took an introductory physics course to ensure that they understood the survey questions as intended. The thinkaloud interviews involved asking each student to answer each survey question while thinking aloud and explain their reasoning. At the end, the interviewer asked further questions if the students did not make their thought processes clear on their own. We found that all of the seven students interpreted all of the survey questions appropriately and there is no qualitative difference between these seven students' responses, who had not taken introductory physics (at the most one or two weeks into an introductory physics course), and the twelve students who were interviewed at the end of an introductory physics course (to be discussed later). Thus, the survey is valid to be administered to the introductory physics students in this investigation both at the beginning and at the end of the course (we also note that a majority of the students had taken at least one high school introductory physics course before the college course).

Moreover, follow-up interviews were conducted with 12 introductory physics students in order to gather qualitative data to complement the quantitative survey data. These follow-up interviews were also conducted using a thinkaloud protocol with each interview lasting for approximately 1 hr. Similar to the interviews at the beginning of the introductory physics course, during the interviews, students first answered the AAPS survey questions along with providing their reasoning for their answers on their own before further questions were asked for clarification of points not made clear. The interviewer then asked questions regarding the way in which students perceived their classes were taught and how they may have shaped their attitudes and approaches to physics problem solving. Thus, these follow-up interviews again served to more deeply probe the students' reasoning behind their responses to the AAPS survey, and to understand their experiences with different methods of instruction. Including the 7 interviews at the beginning of the course and 12 interviews at the end of the course, 11 female and 8 male student volunteers were interviewed (this includes 4 female and 3 male students at the beginning and 7 female and 5 male students at the end of the course).

We made use of the factors identified from a principal component analysis in the original validation study, Ref. [49] to analyze qualitative trends in interviews, e.g., male and female students' responses during the interviews on questions belonging to a particular factor, that will be investigated in detail in future studies. These factors and the questions (items) associated with each one can be seen in Table I. In particular, there were 9 factors identified from a principal component analysis. Factor 1 involves questions related to metacognition and enjoyment of physics problem solving. Factor 2 involves questions about the use of drawing and scratch work while problem solving. Factor 3 involves questions about perception of problem solving approach. Factor 4 involves questions that distinguish between general expert-novice differences in problem solving. Factor 5 involves questions about solving problems symbolically. Factor 6 involves questions about problem solving confidence. Factor 7 involves questions about solving different problems using the same principle. Factor 8 involves questions related to sense making. Finally, Factor 9 involves questions related to problem solving sophistication.

B. Courses and participants

After the matching was done to ensure that each student's responses included both the pre- and post-test data (we note that when initial analysis was conducted without matching students who took both the pre- and post-AAPS survey, no significant differences were found), a total of 678 calculus-based introductory physics students from 8 separate classes from a large state-related research university participated in this study. This group included 422 first semester students (content was mainly mechanics) and 256 second semester students (content was mainly electricity and magnetism). Three sections of first semester physics were taught in a traditional manner and three sections using EBAE methods. The two sections of second semester physics were taught in a traditional manner. With the exception of one section that could have contained possible overlap, the concurrent timing

TABLE I. Principal component analysis results featuring 9 primary factors and description, reproduced from Mason [49]. Note that a given question can be classified into more than one factor.

Factor (% of variance			
explained)	Item	Loading	Description
Factor 1 (13)	13	0.77	Metacognition and
	14	0.73	enjoyment in physics
	7	0.70	problem solving
	10	0.69	
	22	0.63	
	29	0.60	
	20	0.51	
	4	0.57	
	25	0.38	
	27	0.27	
	0	0.28	
	21	0.29	
Factor 2 (8)	18	0.90	Utility of pictures, diagrams,
	17	0.88	or scratch work in physics
	15	0.75	problem solving
	19	0.65	
Factor 3 (6)	5	0.69	Perception of problem
	11	0.68	solving approach
	12	0.55	
	8	0.33	
	26	0.28	
	9	-0.28	
Factor 4 (5)	8	0.76	General expert-novice
	28	0.56	differences in physics
	21	0.43	problem solving
	24	0.43	
	29	0.30	
Factor 5 (5)	31	0.84	Difficulty in solving
	30	0.83	problems symbolically
Factor 6 (5)	1	0.75	Problem solving confidence
	24	-0.71	
	23	0.50	
	6	0.30	
Factor 7 (5)	33	0.88	Solving different problems
	32	0.80	using the same principle
Factor 8 (4)	16	0.76	Sense making
	2	-0.56	C
	5	0.30	
Factor 9 (3)	3	0.77	Problem solving
~ /	20	-0.35	sophistication
	25	-0.32	
	9	0.29	

of when the data were collected enforced that the second semester students were different students from the first semester students, but represented comparable cohorts of students. Because the data were immediately de-identified, it is not possible to know which students may have taken the surveys both in the first and second semesters, but since this overlap was possible only for one section, they would represent a very small percentage of all participants. A majority of students in all these courses are engineering, physical science or mathematics majors and are typically in their first year of college and had taken at least one physics course in high school, although the content and quality of these high school courses can vary greatly.

Traditional classes were operationally defined as classes in which evidence-based active engagement methods were used either very infrequently or not at all. The primary method of instruction for classes defined as traditional was the use of lectures by the course instructor for the majority of instruction. Some classes defined as traditional included limited use of group work, but this use was constrained to recitations or labs run by teaching assistants and only accounted for a small fraction of instructional time. The EBAE classes were operationally defined as classes in which active engagement methods were used frequently. In most cases, this involved course instruction taking place in a fully flipped manner [67,68]. In addition, one class designated as EBAE was not fully flipped but involved a significant amount of Peer Instruction, in which students were frequently engaged with clicker questions and discussion throughout the duration of class time [63], and included frequent group problem solving (at least once per week). No instructor taught both types of classes (i.e., both traditional and EBAE methods).

The average class size was 105 students (an average class size for first semester physics was 86 students, and an average class size for second semester physics was 158 students). Students were free to register for whichever section (of the many sections available) they preferred, and teaching reputations and styles of the different instructors were generally known in advance by the students, or were publicly available online. A summary of the courses can be found in Table II, along with the number of participants included in each class. Because no significant difference was found in initial analysis of unmatched participants, only those who took both the pre- and postsurveys (i.e., "matched") were included in the full analysis. The average loss of student participants due to the missing pre- or postsurvey data was 19%, which includes losses from students who dropped, withdrew, were absent for the pre- or postsurvey, or otherwise chose not to participate in the post-test. Occasionally, some participants were excluded from the matched sets because they added the class late and did not take the presurvey. Because of the many possible reasons for a student missing the pre- and/or postsurvey, it is not possible to infer whether the loss from unmatched to matched is associated with interest level in participation in the surveys.

C. Data collection tools and artifacts

For the first semester physics classes, both the FCI and the AAPS survey were administered twice—once near the TABLE II. A summary of the descriptions of the manner in which each class was taught. We conducted the initial analysis without matching students from pre to post and then performed matched data analysis by matching students from pre to post. We found that the matched analysis was not statistically significantly different from the unmatched analysis so we only discuss matched analysis here. Thus, the numbers listed for matched students (N_{Matched}) are for students included in further analysis. Average loss of students from unmatched to matched in different classes was 19%, and includes losses due to students who dropped or withdrew from the class, were absent for the pre- or postsurvey, e.g., due to adding the class late and missing the presurvey, or choosing not to complete the pre- or postsurvey. The number of matched male and female students is also included. Note that sometimes gender was not indicated by a student. In that case, provided the students had both pre- and postdata, they would have been included in the overall analysis, but not included in the analysis of differences by gender. Thus $N_{\text{male}} + N_{\text{female}}$ may not always equal N_{Matched} .

Class	Description	N _{Matched}	N _{male}	N _{female}	Level
EBAE 1	Flipped classroom instruction	46	31	15	First semester
EBAE 2	Peer instruction or group problem solving	81	54	26	First semester
EBAE 3	Flipped classroom instruction	53	32	20	First semester
Traditional 1	Primarily lecture-based instruction	87	69	18	First semester
Traditional 2	Primarily lecture-based instruction	78	65	13	First semester
Traditional 3	Primarily lecture-based instruction	77	63	11	First semester
Traditional A	Primarily lecture-based instruction	116	90	25	Second semester
Traditional B	Primarily lecture-based instruction	140	105	33	Second semester

beginning of the semester (pre) and once near the end of the semester (post). Similarly, in the second semester classes, both the CSEM and AAPS survey were administered twice (again, once near the beginning and once near the end of the semester). In addition to the FCI and CSEM data, final exam data were collected from all courses. The AAPS survey can be found in the appendix of Ref. [49]. It consists of 33 items (questions) on a five point likert scale spanning "strongly agree," "agree," "neutral," "disagree," and "strongly disagree." For some questions, a favorable or expertlike response to a survey question (based upon faculty responses) is agree or strongly agree, while for other questions a favorable or expertlike response may be disagree or strongly disagree. In addition to the survey data, information was gathered about the teaching methods used, and gender of the students for all classes. We note that the final exams for all physics I courses were common (and similarly, all physics II courses were common) and regardless of the type of instruction, exams for all courses were almost exclusively quantitative problem solving with one or two conceptual questions worth 10% or less of the total credit. Since the common final exam was heavily quantitative, the focus in both traditional and EBAE courses was primarily on helping students develop problem solving skills. The conceptual questions were sometimes used in EBAE classes as clicker questions, but the intent was for students to understand the concepts in order to be able to use them when solving problems.

D. Data collection and analysis methods

Students took the surveys either in lecture or in recitation and were often offered a small amount of bonus points as an incentive for completion. A "dummy" question was included on the survey for all but one class, which was designed to allow the researchers to identify surveys in which the students were not reading the questions carefully, but simply marking down answers randomly. This question asked students who were reading the questions to indicate a specific choice on their bubble sheets for that question. If that choice was not indicated, the students' survey results were not included in the data analysis.

Once all data were collected, normalized scores (which we will simply call scores) for each question were computed by assigning a +1 to a favorable response (regardless of whether it was strongly favorable or favorable), a -1 to an unfavorable response (regardless of whether it was strongly unfavorable or unfavorable), and a 0 to a neutral response. This same convention was used in earlier analysis when the AAPS survey was originally validated [49] and was adopted here for consistency with the validation study. We computed the average normalized score for each question based upon the student responses of +1, 0, or -1, and then averaged these values across all questions on the AAPS to arrive at an average overall normalized score on the AAPS survey for each class.

In our analysis, the AAPS survey data were compiled and examined for changes in scores from pre to post first without matching students (which includes all students who took a pre- or postsurvey regardless of whether they took both surveys) and then again for matched students (which includes only those students who took both the pre- and postsurvey). No significant differences were observed with or without matching students. Therefore, the remainder of the analysis was carried out with matched sets, so that students with a missing presurvey or postsurvey were excluded from the analysis presented here.

III. RESULTS

A. RQ1. How do the pre and post scores on the AAPS survey compare? And how do those measures differ by course (i.e., Physics 1 and Physics 2)?

To answer RQ1, we calculated effect sizes (Cohen's d) [92] to explore differences between the AAPS survey preand postscores (see Table III). We find that, when each section is considered separately, the overall average AAPS survey scores exhibited decline from pre to post, consistent with other types of attitudinal surveys (e.g., those focused on general epistemological beliefs of students), for both first semester and second semester physics instruction as seen in Table III. The average AAPS survey pre- and postscores were also calculated for both first and second semester courses as shown in Table III. We find that when a particular type of course is averaged, the difference between the pre and postscores is statistically significant only for the traditional courses (see Table III). Table III also displays the small effect sizes for the declines. Moreover, the presurvey scores were moderately correlated with the postsurvey scores on the AAPS survey (with correlation coefficient $R \approx 0.5$).

As can be seen in Table III, the AAPS survey average scores were similar in the second semester compared with the first semester. The second semester classes in this investigation were all instructed in a traditional manner. We find that the average pre AAPS survey scores were between approximately 0.50 and 0.55 for both first and second semester students, and post were between approximately 0.40 and 0.50 for both first and second semester students. Although the majority of data from the second semester do not include the same individual students as those in the first semester of introductory physics, the cohorts of students in the first semester are comparable to those in the second semester. Some possibilities for comparable prescores for the two semesters are that students' attitudes may rebound when they begin their second semester and/or those who

TABLE III. p values obtained via the t tests comparing the average pre- and post-AAPS survey scores, and effect sizes for the decline in scores from the pre- to post-AAPS survey.

Class	Sem	Pre	Post	<i>p</i> value pre vs post	Effect size
EBAE 1	1	0.555	0.518	0.099	-0.16
EBAE 2	1	0.547	0.507	0.043	-0.21
EBAE 3	1	0.557	0.496	0.017	-0.28
Avg. EBAE	1	0.552	0.507	0.099	-0.12
Traditional 1	1	0.517	0.447	0.005	-0.39
Traditional 2	1	0.465	0.436	0.146	-0.15
Traditional 3	1	0.497	0.423	< 0.001	-0.34
Avg. Trad	1	0.494	0.434	0.004	-0.17
Traditional A	2	0.508	0.402	< 0.001	-0.49
Traditional B	2	0.552	0.453	< 0.001	-0.44
Avg. Trad	2	0.532	0.431	< 0.001	-0.33

continue to the second semester have better attitudes at the beginning of the course than those in classes taught traditionally in the first semester. In particular, it is possible that some students may not continue into the second semester of physics and these students may be the ones with less favorable attitudes. Yet another possibility is that, since the second semester physics courses were all taught in a traditional manner, students from first semester EBAE courses who have more favorable average attitudes and approaches (who continue into the second semester course) could inflate the presurvey scores in the second semester traditionally taught course. However, regardless of the reason for the increase at the beginning of the second semester (compared to the end of the first semester), by the end of the second semester, the average score was similar to that which was observed at the end of the first semester.

B. RQ2. How do AAPS survey scores in physics 1 compare for EBAE vs traditional instruction without separating students by gender and how do they compare when male and female students are considered separately?

To answer RQ2, an ANCOVA [93] analysis was performed and the results are discussed below.

EBAE vs traditional instruction: We find that, on average, students in the EBAE classes scored higher than those in the traditional classes on the AAPS survey before instruction, as can be seen in Fig. 1. The difference in presurvey scores between the traditional and EBAE classes was statistically significant (p = 0.004) suggesting that on average students who had a more positive attitude about problem solving and/or EBAE methods may have chosen instructors who use the EBAE method of instruction over those who use traditional instruction when they enrolled (although different sections of the course were not listed in the physics course schedule as EBAE or traditional, students had knowledge of which section of the course



FIG. 1. Raw normalized data (i.e., without controlling for presurvey scores) show the overall average scores for both the pre- and post-AAPS survey in the calculus-based first semester classes. The error bars represent standard error.



FIG. 2. Controlling for the AAPS survey pretest score in first semester classes for method of instruction, the AAPS post-test scores for traditional classes is significantly lower than that of EBAE classes [F(2, 418) = 76.498, p < 0.001]. Error bars represent standard error.

would be taught by whom, and this information along with each instructor's teaching approach available online can potentially be used to select a particular instructor).

To examine whether there are differences in the post-test AAPS survey scores after controlling for the pretest scores for the EBAE and traditional methods of instruction, we performed an ANCOVA [93] in which the covariant is the pretest scores, the dependent variable is the average AAPS survey post-test score, and the independent variable is the method of instruction (i.e., EBAE vs traditional). After controlling for the pretest score for the different methods of instruction, we find a statistically significant difference in post-test score by method of instruction [F(2,418) =76.498, p < 0.001 [93], with the average AAPS survey postscore being higher for the EBAE classes. The adjusted post-test scores, after controlling for the pretest scores for the different methods of instruction, appears in Fig. 2. In particular, the adjusted post-test score for traditional classes, after controlling for the pretest score, is 0.467 with a 95% confidence interval of [0.439,0.495], and the adjusted AAPS survey post-test score for EBAE classes, after controlling for the pretest score, is 0.511 with a 95% confidence interval of [0.481,0.540] [93].

Interviews hint at the fact that students in the EBAE classes may have more expertlike attitudes about factor 1 focusing on metacognition in problem solving after a semester of instruction compared with their traditionally taught peers. For example, one EBAE student who answered favorably to the statement, "When I solve physics problems, I always explicitly think about the concepts that underlie the problem," went on to emphasize how important thinking about the concepts is to problem solving. He stated, "I think you need to know those concepts to understand what's going on... I think the more conceptual understanding that's there, the better." Further discussion

suggests that this EBAE student had an expertlike perspective on the importance of carefully thinking about the concepts during problem solving and that his class activities reinforced these views. In interviews, some students from classes instructed in an EBAE manner verbalized their reasoning regarding the questions categorized into factor 4 that demonstrated expertlike attitudes and approaches to problem solving (this factor includes awareness of more than one way to solve a problem, applying the same principle to different contexts, and perseverance in problem solving). For example, regarding the statement "After I have solved several physics problems in which the same principle is applied in different contexts, I should be able to apply the same principle in other situations," one student from an EBAE class stated, "We did force and Newton's laws, and then we kind of apply those same principles now, so it's the same way of going about the problem." Further discussions suggest that this student in the EBAE course felt that his course and solving problems in small groups helped him discern the connections between problems solved earlier in the course and the problems he encountered later, in terms of how the same principles apply, suggesting a more expertlike attitude about underlying principles when solving problems. These qualitative findings from interviews will be investigated in detail in the future.

Gender differences: When we separated our data by gender and analyzed the responses, we find that unlike gender differences in content-based surveys discussed earlier, there is often a gender gap in which female students' scores on the AAPS survey are *higher* than those of male students [56,57]. Not only were average AAPS survey scores for female students often higher than those for male students, they also remained higher at the end of instruction. Figures 3–6 show that female students exhibited less of a



FIG. 3. Raw normalized data (i.e., without controlling for prescores) show the first semester average AAPS scores by gender and method of instruction. Average male scores (M) are indicated by squares for traditional male students (total number n = 197) and open circles for EBAE male students (n = 118). Average female scores (F) are indicated by diamonds for traditional female students (n = 42) and closed circles for EBAE female students (n = 61). Error bars represent standard error.



FIG. 4. After controlling for the pretest scores by both gender and method of instruction in first semester classes, there is a statistically significant difference in post-test score by gender with female students scoring significantly higher than male students and by method of instruction with students in the EBAE courses scoring significantly higher than those in traditionally taught classes [F(3, 413) = 54.575, p < 0.001]. We note that there was no interaction between gender and method of instruction (p = 0.782). Error bars represent standard error.

decline on the AAPS survey pre to post compared to male students regardless of whether they were instructed traditionally or with EBAE methods. Since the decline among female students was negligible in the first semester, this suggests that the overall decline on the AAPS survey in the course was mainly driven by the male students because male students outnumbered female students (as can be see in Table II).

To examine whether there are differences in the post-test AAPS survey scores after controlling for the pretest scores for both gender and method of instruction, we performed an ANCOVA [93] in which the covariant is the pretest scores, the dependent variable is the average AAPS survey



FIG. 5. Raw normalized data (i.e., without controlling for prescores) show the second semester AAPS survey scores by gender. Average male scores (M) are indicated by squares (n = 195) and average female scores (F) are indicated by diamonds (n = 58). Error bars represent standard error.



FIG. 6. After controlling for the AAPS survey pretest score in second semester classes for gender, there are statistically significant differences in AAPS survey post-test scores with female students outperforming male students [F(2, 250) = 59.867, p < 0.001]. Error bars represent standard error.

post-test score, and the independent variables are method of instruction and gender. In particular, controlling for the pretest scores by both gender and method of instruction in first semester classes, we find that there is a statistically significant difference in post-test score by gender with female students scoring significantly higher than male students and by method of instruction with students in the EBAE courses scoring significantly higher than those in traditionally taught classes [F(3, 413) = 54.575], p < 0.001]. We note that there was no interaction between gender and method of instruction (p = 0.782). The adjusted post-test scores for the first semester classes, after controlling for the pretest scores, appear in Fig. 4. We note that after controlling for pretest score by gender and method of instruction, the adjusted average AAPS survey post-test score for male students in traditional classes is 0.442 with a 95% confidence interval of [0.415, 0.469], the adjusted AAPS survey post-test score for female students in traditional classes is 0.496 with a 95% confidence interval of [0.438,0.555], the adjusted AAPS survey post-test score for male students in the EBAE classes is 0.489 with a 95% confidence interval of [0.454,0.524], and the adjusted AAPS survey post-test score for female students in the EBAE classes is 0.531 with a 95% confidence interval of [0.482,0.579] [93].

However unlike first semester classes, female students' attitudes and approaches to problem solving measured by the AAPS survey appear to decline more in second semester classes (see Fig. 5). Nevertheless, when pretest scores are controlled for gender, there are statistically significant differences by gender in the second semester classes, with female students outperforming male students [F(2, 250) = 59.867, p < 0.001] [93]. The adjusted posttest scores for first semester classes, after controlling the pretest score for gender appear in Fig. 6. In particular, after controlling the pretest score for the gender of students, the adjusted AAPS survey post-test score for male students in

the second semester is 0.414 with a 95% confidence interval of [0.386, 0.442], while the adjusted AAPS survey post-test score for female students in the second semester is 0.483 with a 95% confidence interval of [0.431,0.534] [93].

Interviews revealed further gender differences in attitudes and approaches to problem solving, particularly when it came to working with other students. Female students who were interviewed often reported feeling less comfortable and/or confident when working with male peers; whereas all interviewed male students reported feeling comfortable with either gender. Some reasons female students felt less comfortable working with male peers were revealed in interviews. For example, one female student reported that she feels discouraged when she works with male peers, stating: "It's usually guys that say you're totally wrong, but they don't always know themselves, so I think it's unfortunate that their first response is to say that." Conveying a similar sentiment, another female student described her interactions with male peers in the class as follows, "I feel like the majority of the time, if a guy knows how to do a physics problem he will let you know he knows how to do a physics problem." This female student emphasized that she observes that male students confidently assert what they know. Another female student explained that she felt intimidated by working with male students, stating: "I think I go to females first. I think I get more intimidated if I ask a guy because I think they're a lot more condescending about it and I don't like it." Further discussions suggest that many female students felt that their male peers may sometimes be dominant and/or overconfident, particularly in their interactions with female students and such interactions made them less sure about themselves so they would rather not interact with them if they can avoid it.

Moreover, hints of higher level of confidence came through in some interviews with male students. For example, one male student described his approach to learning from problem solving as "the best approach" stating "start with simple problems and then work your way up but if you really want to understand a physics concept you should be doing the hard problems and you will understand it deeper." Further discussions suggest that he had a high level of confidence that his approach was better than others'. In contrast, a female student who described a somewhat similar approach did not claim that it was the best approach and stated, "...if you have a bunch of easy problems, it's not going to help you in the long run in a similar scenario because you won't have been able to work out the harder problem." Another male student, when discussing whether he would trace back his work for mistakes stated that he does not generally do it because, "It's really hard to find mistakes because it's my work. To me it looks perfect." In general, male students interviewed were more confident and sometimes explicitly labeled their approaches to problem solving as the "best"

or their problem solution as "looking perfect" to them, something that none of the interviewed women ever mentioned. In fact, female students who were interviewed never spoke with such high levels of confidence about physics problem solving and they often harbored doubts about themselves. For example, one female student, while discussing the use of gut feeling and intuition in physics problem solving said, "My gut is always wrong about everything." This female student further noted that she never uses her intuition since it is always wrong. Another female student also noted that she did not feel confident using her gut feeling "because my gut differs from the rules I've been taught." The same female student noted that she prefers to follow rules in physics problem solving rather than follow her gut feeling. She also reflected upon her feelings when solving challenging physics problems stating: "They can make you feel defeated or they can make you feel really proud of yourself." However, she noted that she found solving challenging problems to be stressful because she was not sure that she would be able to solve them correctly. Another female student, discussing the role of making approximations in physics, said that she was always worried about whether it was appropriate to make a certain approximation in a given case stating "...you can't assume everything because while a lot of physics is what happens around you, a lot is unseen. Like when I draw free body diagrams-it would often trip me up because you want to assume the simpler case but you can't always..." Overall, while AAPS survey scores are higher for female students suggesting better attitudes and approaches to problem solving, interviewed female students, in general, expressed lower levels of confidence while solving physics problems compared to the male students.

One hypothesis is that the apparent difference in confidence level based upon gender observed in the interviews may be connected to the way in which male and female students responded to questions on the AAPS related to approaches to problem solving. In particular, it is possible that the lower confidence levels of female students could motivate them to use more careful and conscientious problem solving practices. For example, interviewed female students often reported the consistent and frequent use of scratch work and drawings as a tool for helping them carefully sort out how to solve a problem (factor 2). One female student described drawing a picture or diagram for each and every problem to make sure she was on the right track, stating "Even if every picture looks the same, I just have to draw it." Another female student describing her use of scratch work stated, "I need something to keep my mind in one place. If I don't draw and try to free-think it, I will overcomplicate it." This female student further added that drawing and scratch work helps her carefully organize her thought processes in order to increase the probability of success in physics problem solving. In interviews, doing scratch work and drawing diagram are two very tangible examples that female students emphasized resorting to during physics problem solving much more than male students regardless of whether the problem was a multiplechoice problem or open-ended problem and valuing of these effective strategies potentially hints at more careful problem-solving practices among many female students. In particular, interviewed male students were less likely to emphasize the importance of consistent use of drawings and scratch work in physics problem solving, especially for multiple-choice problems. For example, in an interview, one male student said that he does not focus much on scratch work or diagram in solving multiple-choice problems since there was no credit associated with the process of showing the work. He added, "My work is always scattered around...if it's multiple choice." On the other hand, most interviewed female students indicated that they would even do scratch work or drawings for multiple choice questions. For example, one female student in an interview explained, "I like to do scratch work no matter what...I'll always draw a picture, because even for multiple choice it's just as important because it helps yield a correct answer." Similarly, another female student stated, "Scratch work is always helpful... It can help you understand what's going on." Other female students interviewed were also more likely to indicate than male students who were interviewed that scratch work and drawing are crucial in solving a physics problem.

Other hints of valuing of more careful problem solving approaches on the part of female students were found in interview responses which often conveyed a more reflective overall approach. One female student described how her process of solving problems often starts: "...my method of answering physics problems-what are they asking me about?" This student further described how she carefully reflects on what the problem is asking her before proceeding and always draws diagrams as appropriate. Another female student described reflecting upon different approaches and reconciling any inconsistencies stating, "If two equations give me two different answers, it will drive me crazy until I figure out which one is wrong." This female student further expressed that in her opinion it is extremely important to reflect upon the different approaches used to solve a problem to understand which one is correct. When thinking about the correctness of her answers, another female student recalled an example when, "I got a really small acceleration that made me think, that's not possible," and how it was very important for her to spend time figuring out what went wrong. Another female student who had earlier described her frequent use of drawings stated, "I always check my work and see if my picture makes sense and if I made a mistake there." These female students suggested that part of their problem solving strategy was to reflect upon the soundness of their approaches and answers. In addition, more than male students, interviewed female students often discussed at length the importance of reflecting on real life

applications of physics in order to solidify concepts. For example, one female student said, "Physics is about the world around you. You aren't just doing numbers; You're comparing it to a real life scenario." Another female student enthusiastically stated, "It's so nerdy but I'll come out of lecture and I'll see a leaf falling and think about the physics. I can't see people run without thinking, 'I wonder how fast they're going,' or a frisbee being thrown and think 'I wonder what angle that was thrown at'." Further discussions suggest that these real-life connections were very important part of learning physics for her.

C. RQ3. How correlated are students' AAPS survey scores in a given class with their conceptual survey (FCI or CSEM) or final exam performance?

To answer RQ3, the correlation coefficients between the AAPS survey scores and the FCI, CSEM, or final exam performance were calculated for male and female students in both the EBAE and traditionally taught courses. However, before examining possible correlation between AAPS survey score and conceptual survey or final exam performance, it is helpful to first put the conceptual survey and final exam scores into perspective.

As seen in Figs. 7 and 8, the FCI and CSEM performances were similar to typical values in other studies at other institutions [38,46,47,56,68]. However, the normalized gain for the FCI was statistically significantly higher (p = 0.033, d = 0.169) for EBAE classes compared with traditional classes. The FCI scores exhibited a gender gap which was present in both EBAE and traditional classes. For traditionally instructed first semester students, the gender gap was



FIG. 7. Average FCI scores by gender and type of instruction. Average male student scores (*M*) are indicated by squares for traditionally taught male students (total number n = 197) and open circles for the EBAE male students (n = 118). Average female student scores (*F*) are indicated by diamonds for traditionally taught female students (n = 42) and closed circles for EBAE female students (n = 61). A gender gap that is significant is seen for both the traditional instruction and the EBAE instruction. Normalized gain for female students in EBAE classes compared with female students in traditional classes is statistically significant (p < 0.001).



FIG. 8. Second semester (traditional instruction) average CSEM scores show a gender gap which is not initially statistically significant (p = 0.060) in presurvey scores but becomes statistically significant in postsurvey scores (p = 0.017). The total number n = 195 for male students (M) and n = 58 for female students (F).

statistically significant, of large effect size (p < 0.001,d = 0.86) for the average presurvey score and statistically significant, of large effect size (p < 0.001, d = 0.98) for the average post survey scores. Similarly, for first semester students who received EBAE methods of instruction, the gap was statistically significant, of large effect size for average presurvey scores (p < 0.001, d = 0.93), but dropped to medium effect size for postsurvey scores (p < 0.001, d = 0.64). The gap did not decrease or increase statistically for either EBAE classes or traditionally taught classes, based upon comparison of male and female students' normalized gain. However, the difference in normalized gain for female students in EBAE classes compared to female students in traditionally taught classes was statistically significant, of medium effect size (p < 0.001, d = 0.63). This suggests that female students may have gained more conceptual understanding in EBAE classes compared with traditional classes in their first semester of instruction. For second semester students, there was a gender gap in CSEM performance that was not statistically significant for pretest scores (p = 0.060), but was statistically significant in the post-test scores, of small effect size (p = 0.017, d = 0.30). There was no statistically significant change in the gender gap based on comparison of normalized gain for male vs female students in the second semester course.

When the final exam scores were examined for differences by method of instruction for first semester classes, we find that there is a statistically significant difference in the final exam scores based upon the method of instruction. It is important to note that the final exams for all sections were common regardless of the type of instruction, so that the final exam performance can be appropriately compared. The EBAE classes scored an average of 70% on the final exam, which was significantly

higher, of small effect size (p < 0.001, d = 0.43) than the average score in traditional classes, which was 62%. On the other hand, no statistically significant difference was found in final exam score by gender in first semester classes, either in EBAE classes (p = 0.166) or traditional classes (p = 0.207). Likewise, there was no statistically significant difference by gender in second semester classes (average score was 60%, p = 0.415). These findings agree with previous studies that have shown small or no gender difference in final exam scores, even when there is a gender gap on conceptual survey performance [72,94].

When we examined the final exam and FCI or CSEM scores for possible correlation with the AAPS survey scores, we find that the overall AAPS survey scores and FCI or CSEM scores were not well correlated. All sections have correlation coefficients of R < 0.30 when considering all students.

However, some small to medium correlations were found when separating the data by gender in first semester classes as can be seen in Table IV (for second semester classes the correlations by gender were weaker). In particular, performance on the FCI post-test and on final exams is correlated with the AAPS survey post-test performance for female students in EBAE classes with a medium correlation coefficient (R = 0.40-0.46). Also, as seen in Table IV, the final exam performance for female students in first semester traditional classes is correlated with the AAPS survey post-test score with medium correlation coefficient (R = 0.48). Table IV suggests that the performance on content is more correlated with AAPS scores for female students. One hypothesis for why in Table IV, for female students, the correlation coefficient between the FCI and AAPS (0.21) is lower than the correlation coefficient between the final exam and AAPS (0.48) in the traditional courses but these correlation coefficients are comparable (0.40 and 0.46) in the EBAE courses is that in EBAE courses there was more emphasis on conceptual understanding and using it to solve problems (as opposed to in the traditional courses in which conceptual understanding is generally underemphasized).

In addition, since we saw a correlation in the first semester classes by gender, we did a *post hoc* analysis

TABLE IV. Correlation coefficients (*R*) for the AAPS survey scores and final exam or FCI scores for different types of courses in first semester classes, broken down by gender. R > 0.3 appears in boldface.

		<i>R</i> for AAPS post and	<i>R</i> for AAPS post and
Class	Gender	FCI post	final
Traditional 1st semester	Μ	0.24	0.22
Traditional 1st semester	F	0.21	0.48
EBAE 1st semester	Μ	0.13	0.11
EBAE 1st semester	F	0.40	0.46

TABLE V. Correlation coefficients (*R*) for post-test scores for the AAPS survey question 16 (which asks whether students use physics principles or "gut" feeling when answering conceptual questions) and the FCI for different types of courses in first semester classes. R > 0.3 appears in boldface.

Class	Gender	<i>R</i> for AAPS question 16 and FCI post
Traditional 1st semester	М	0.06
Traditional 1st semester	F	0.27
EBAE 1st semester	М	0.02
EBAE 1st semester	F	0.47

of the correlation between question 16 on the AAPS survey and FCI scores in first semester classes. This question (question 16) states: "When answering conceptual physics questions, I mostly use my 'gut' feeling rather than using the physics principles I usually think about when solving quantitative problems." A favorable response would be to disagree with this statement. One motivation to explore this particular question for possible correlation with the FCI scores was that there appear to be possible differences via response to this AAPS question by method of instruction. When correlation was examined for this question with FCI score, it was found that there was a small positive correlation for female students between their score on this question in the post-test and their overall post-test FCI score. In particular, for female students in traditionally taught classes, the correlation coefficient was found to be R = 0.27, and for female students in EBAE classes, the correlation coefficient was found to be R = 0.47. These correlation coefficients are summarized in Table V. The higher correlation between question 16 response and performance on conceptual survey performance, particularly for female students, is interesting and will be explored further in future investigations.

IV. DISCUSSION

A. Findings regarding EBAE compared to traditional instruction

Our findings suggest that after controlling for the pretest scores for method of instruction, EBAE students' attitudes and approaches to problem solving remain more favorable at the end of a semester of instruction, as compared to traditionally taught students. Even though the EBAE classes started out with higher scores on their AAPS presurvey, the decline they exhibited from the pre- to postsurvey in EBAE classes was found to be statistically significantly lower than the decline in similar traditional classes.

Furthermore, we note that some students may be "selfselecting" by enrolling into sections taught by instructors who use EBAE methods. In particular, instructors who teach the EBAE classes may be attracting students who may have more positive attitudes towards problem solving compared with students who enroll in traditionally taught classes, as evidenced by the small but statistically significant difference in presurvey scores on the AAPS survey, favoring EBAE students. Better understanding of the reasons for such an offset, and uncovering whether or not other universities observe a similar trend in their large enrollment introductory physics courses (in which several sections of the same course are taught simultaneously) would be an interesting future research avenue to pursue. In addition, our findings provide some evidence that EBAE instruction may be beneficial for students in learning physics concepts, since both the FCI normalized gain and the final exam scores were statistically significantly higher for the EBAE classes compared with traditionally taught classes.

B. Findings related to gender

Our findings indicate that the post AAPS survey scores were higher and more consistent from the beginning to the end of the semester for female students than for male students. Female students exhibited almost no decline in the AAPS survey score from pre- to postsurvey in first semester physics and less of a decline than male students in second semester physics. Moreover, correcting for pretest scores, significant differences in the AAPS survey post-test scores were found by gender, as seen in Fig. 4. Indeed, it appears that male students, who outnumber female students significantly, are driving the overall decline in AAPS survey scores from the pre- to postsurvey. This is somewhat unlike other attitudinal surveys (e.g., those focused on epistemological beliefs about physics) in which female students, on average, show more (or at least as much) of a decline in scores from pre- to postsurvey [38,47]. Because the AAPS survey incorporates questions related to attitudes and approaches to problem solving, not just epistemological beliefs of students about physics and physics learning like other surveys, it is possible that female students may be answering these questions (particularly those related to approaches to problem solving) differently than questions in previous surveys. The factor related to diagrams and scratch work is an example of a cluster on which interviewed female students in general tended to provide qualitatively different responses. Future investigation would probe this trend further.

C. Connections between gender, attitude, and performance

Our findings suggest that female introductory physics students, on average, have attitudes and approaches to problem solving that are more expertlike than their male peers, especially at the end of the course. Moreover, female students' attitudes and approaches do not suffer the same decline that male students' do. These findings are novel and different from those found using broad epistemological surveys. They suggest that instructional design should incorporate strategies to take advantage of female students' positive attitudes and approaches when it comes to solving physics problems and their persistence in maintaining more expertlike attitudes and approaches from the beginning to the end of the semester.

Yet it is equally well known that women are significantly underrepresented in calculus-based introductory physics courses discussed here and in physics overall. Introductory physics classes give students their first experience with physics, and could open doors to further pursuit of physics. Although gender gaps, in which female students' performance lags behind that of male students, are known to exist in tests of conceptual knowledge and/or problem solving, the correlations we have found among female students between their attitude and approach to problem solving and scores on conceptual surveys and exams suggest a possible path forward. If female students' attitudes and approaches to problem solving show promise, as we have found, we may be able to take advantage of this finding to support their learning in introductory physics classes.

Also, the fact that attitudes and approaches to problem solving appear to have higher correlation to conceptual and/or final exam performance for female students may suggest the possibility that female students' performance may benefit more from improved attitudes and approaches to physics problem solving. One hypothesis is that the higher correlation between attitudes and approaches to problem solving and their actual performance for female students may in part be related to the implicit stereotype threat [95–97] they experience simply by being in a physics class in which they are severely underrepresented. In particular, since stereotype threat can consume some cognitive resources from the limited capacity of working memory for female students, it can leave them with fewer cognitive resources available for tasks related to physics problem solving. However, more expertlike attitudes and approaches to problem solving may aid in keeping cognitive load [80] under check despite the stereotype threat that many female students automatically experience in physics problem solving [77]. Thus, more expertlike attitudes and approaches to problem solving may be better correlated with their physics performance than for male students. Furthermore, this possible explanation suggests that the gender performance gap may be reduced by encouraging favorable attitudes and approaches towards problem solving for all students including female students (whose AAPS scores are more correlated with physics performance).

In addition to a higher correlation between attitude and approach to problem solving and conceptual performance, we also found that with regard to approaches to problem solving, female students valued engaging in the use of drawings and scratch work more than male students. We hypothesize that these findings may be related. In particular, it may be possible to take advantage of the fact that more than male students, female students were found to value drawings and scratch work while solving physics problems to promote conceptual understanding. For example, one interviewed female student explained why she draws diagrams whenever possible as follows: "I tend to draw a picture or diagram to help me visualize and understand conceptually what is going on." She added that this step is very important for her to interpret the physics problems appropriately. Responses from female students such as these suggest a possible strategy that may support learning physics is to leverage their better attitudes and approaches to problem solving using pedagogical interventions and this could even serve to reduce the gender performance gap that is often seen in conceptual surveys discussed earlier. This hypothesis will be tested in future investigations.

Moreover, at least part of the success of these EBAE methods, particularly for female students, may lie in the extent to which instructors address their sense of belonging, self-efficacy and mindset [98–100] in a physics class in which they are often underrepresented [68,83–90]. Although we did not investigate these issues, interviews suggested that female students may sometimes feel less comfortable working with male students in an EBAE class. These issues may be related to the fact that the self-efficacy and sense of belonging of female students in a physics class in which they are outnumbered by male students can be negatively impacted by such interactions in the EBAE courses if these motivational issues are not accounted for explicitly in the design of the EBAE courses.

If female students' sense of belonging, self-efficacy, and mindset can be positively impacted through course design and/or the use of mindset interventions [98–100], then it may be possible to powerfully leverage the positive attitudes and approaches of the female students to improve learning.

An encouraging theme that arose in interviews was the tendency of many female students to express what appeared to be expertlike sense making practices. For example, one interviewed female student stated that while solving physics problems, "It's very important to be able to see [visualize] it and understand it. It's not just numbers or it would be a math class... equations are not things you should memorize for convenience. They should mean something." This student further described the importance of a deliberate effort to make sense of what was going on in the given problem, emphasizing that physics is not simply equations devoid of meaning. Another female student said that "If I don't think about the concepts, then the math doesn't make sense to me" and went on to describe her systematic sense making approach as a way to solve challenging physics problems. Sentiments such as these suggest that one fruitful path to explore in future research is whether women in these types of calculus-based introductory physics courses are more engaged in sense making than their male counterparts, on average, while solving physics problems. These hypotheses will be investigated in the future.

V. SUMMARY

We investigated the impact of instruction on attitudes and approaches to problem solving as measured by the AAPS survey administered both at the beginning and end of semester long introductory physics courses at a large research university in the United States. We also compared the AAPS survey scores based upon method of instruction and gender of students, and examined data for correlations between the AAPS survey scores and performance on conceptual surveys (FCI and CSEM) or final exams. Moreover, after controlling for the pretest scores by method of instruction, the use of evidence-based active engagement methods had a positive impact on student attitudes and approaches to problem solving. Furthermore, there was a decline in students' scores on the AAPS survey from the beginning to the end of the semester, but female students generally showed less decline than male students. In addition, controlling for the AAPS pretest scores, the AAPS posttest scores were statistically significantly higher for female students compared to make students for both types of instruction (EBAE or traditional lecture-based). The correlation between the AAPS survey scores and performance on conceptual surveys or final exams appeared stronger for female students compared with male students, especially in the EBAE courses. Similarly, correlations were higher for individual AAPS survey questions analyzed and FCI score for female students than for male students. Finally, although data are not presented here, we found that students at a different large research university exhibited similar scores on the AAPS survey in similar types of classes. Thus, these findings are likely to be generalizable at least to other large research universities.

VI. IMPLICATIONS

These findings can be useful for instructors concerned with the type of instructional method that may promote favorable attitudes and approaches to physics problem solving among their students. In addition, those concerned with the under-representation of women in physics may find these findings illuminating and take advantage of them in their instructional design, given the promising responses that female students provide on the AAPS survey. For example, since our findings suggest that students in the EBAE courses had better attitudes and approaches to problem solving as well as better physics scores, instructors should employ EBAE approaches in their courses. Equally importantly, since compared to male students, female students had significantly better attitudes and approaches to problem solving at the end of the course and female students' AAPS scores were correlated with their scores on

FCI and also correlated with their final exam scores in the EBAE courses, use of pedagogies that promote effective attitudes and approaches to problem solving has the potential to bridge the gender performance gap. Many prior investigations [1-37] have focused on strategies to improve students' approaches to physics problem solving by providing different types of scaffolding support and incentives. However, these studies have not explicitly focused on how these pedagogies striving to improve students' approaches to problem solving impacted male and female students' attitudes and approaches and whether the gender gap in performance was reduced using these pedagogies focusing explicitly on effective approaches to problem solving. One hypothesis for why more expertlike attitudes and approaches to problem solving may be better correlated with the physics performance for female students and instruction promoting and incentivizing effective problem solving strategies throughout their courses has the potential to reduce the gender performance gap is that many female students are likely to be victims of stereotype threat in physics courses due to societal biases. The stereotype threat can consume some of their limited cognitive resources while solving physics problems. However, effective approaches to problem solving, e.g., performing a conceptual analysis of the problem and drawing a diagram, writing knowns and unknowns, and decomposing the problem into subproblem can distribute the cognition [101] between a student's working memory and whatever the student writes as part of the systematic solution. Therefore, sufficient cognitive resources are more likely to be available for metacognition at a given time during problem solving despite the fact that some cognitive resources are taken up by the stereotype threat. Also, if female students' valuing of sense making found in this research during interviews can be leveraged by instructors to help them engage deeply in sense making when solving physics problems (in a supportive environment which does not cause increased anxiety), it may be possible to improve their learning significantly.

ACKNOWLEDGMENTS

We thank the NSF for Grant No. DUE-1524575 and Sloan Foundation for Grant No. G-2018-11183. We are very grateful to Dr. Emily Smith and Professor Leon Hsu for providing data from one university and for helpful discussions and to all the faculty at the other university and to all students who helped with this investigation. We thank Yasemin Kalender and Professors Tim Nokes-Malach, Chris Schunn, Suzanne Lane, and Sarai Hedges for all their help with statistical analysis and discussions about the AAPS survey (including their time in reading the earlier research study involving the validation of the survey).

- [1] J. Docktor and P. Mestre, Synthesis of discipline-based educational research in physics, Phys. Rev. ST Phys. Educ. Res. **10**, 020119 (2014).
- [2] D. Maloney, An Overview of Physics Education Research on Problem Solving, in Getting started in PER, Reviews in PER, 2 (AAPT, 2011), https://www.compadre.org/per/ items/detail.cfm?ID=11457.
- [3] M. Chi, P. Feltovich, and R. Glaser, Characterization and representation of physics problems by experts and novices, Cogn. Sci. 5, 121 (1981).
- [4] B. Eylon and F. Reif, Effects of knowledge organization on task performance, Cognit. Instr. 1, 5 (1984).
- [5] F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. 63, 17 (1995).
- [6] J. I. Heller and F. Reif, Prescribing effective human problem solving processes: Problem description in physics, Cognit. Instr. 1, 177 (1984).
- [7] P. Hardiman, R. Dufresne, and J. Mestre, The relation between problem categorization and problem solving among novices and experts, Memory Cognit. 17, 627 (1989).
- [8] J. Mestre, R. Dufresne, W. Gerace, P. Hardiman, and J. Touger, Promoting skilled problem solving behavior among beginning physics students, J. Res. Sci. Teach. 30, 303 (1993).
- [9] W. Leonard, R. Dufresne, and J. Mestre, Using qualitative problem solving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. 64, 1495 (1996).
- [10] R. Dufresne, W. Gerace, P. Hardiman, and J. Mestre, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, J. Learn. Sci. 2, 307 (1992).
- [11] J. Mestre, Probing adults' conceptual understanding and transfer of learning via problem posing, J. Appl. Dev. Psychol. 23, 9 (2002).
- [12] R. Dufresne, J. Mestre, T. Thaden-Koch, W. Gerace, and W. Leonard, Knowledge representation and coordination in the transfer process, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 538–553.
- [13] J. Mestre, Is transfer ubiquitous or rare: New paradigms for studying transfer, AIP Conf. Proc. 790, 3 (2005).
- [14] T. Nokes-Malach and J. Mestre, Toward a model of transfer as sense-making, Educ. Psychol. 48, 184 (2013).
- [15] K. Harper, Student problem solving behaviors, Phys. Teach. 44, 250 (2006).
- [16] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, Am. J. Phys. 60, 627 (1992).
- [17] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, Am. J. Phys. 60, 637 (1992).
- [18] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving I. Mapping the common core, Phys. Rev. ST Phys. Educ. Res. 3, 020109 (2007).

- [19] D. Meltzer, Relation between students' problem solving performance and representational mode, Am. J. Phys. 73, 463 (2005).
- [20] K. K. Mashood and V. Singh, Large-scale studies on the transferability of general problem solving skills and the pedagogic potential of physics, Phys. Educ. 48, 629 (2013).
- [21] G. Gladding, B. Gutmann, N. Schroeder, and T. Stelzer, Clinical study of student learning using mastery style versus immediate feedback online activities, Phys. Rev. ST Phys. Educ. Res. 11, 010114 (2015).
- [22] L. Ding, N. Reay, A. Lee, and L. Bao, Using conceptual scaffolding to foster effective problem solving, AIP Conf. Proc. **1179**, 129 (2009).
- [23] C. Singh, When physical intuition fails, Am. J. Phys. 70, 1103 (2002).
- [24] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. I. Performance on a nonintuitive problem pair from introductory physics, Phys. Rev. ST Phys. Educ. Res. 4, 010104 (2008).
- [25] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer, Phys. Rev. ST Phys. Educ. Res. 4, 010105 (2008).
- [26] C. Singh, Problem solving and learning, AIP Conf. Proc. 1140, 183 (2009).
- [27] S. Y. Lin and C. Singh, Challenges in using analogies, Phys. Teach. 49, 512 (2011).
- [28] S. Y. Lin and C. Singh, Using isomorphic problems to learn introductory physics, Phys. Rev. ST Phys. Educ. Res. 7, 020104 (2011).
- [29] S. Y. Lin and C. Singh, Using isomorphic problem pair to learn introductory physics: Transferring from a two-step problem to a three-step problem, Phys. Rev. ST Phys. Educ. Res. 9, 020114 (2013).
- [30] S. Y. Lin and C. Singh, Effect of scaffolding on helping introductory physics students solve quantitative problems involving strong alternative conceptions, Phys. Rev. ST Phys. Educ. Res. 11, 020105 (2015).
- [31] E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context, Phys. Rev. ST Phys. Educ. Res. 8, 020109 (2012).
- [32] E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, What do students do when asked to diagnose their mistakes? Does it help them? II. A more typical quiz context, Phys. Rev. ST Phys. Educ. Res. 8, 020110 (2012).
- [33] A. Mason, E.Yerushalmi, E. Cohen, and C. Singh, Learning from mistakes: The effect of students' written self-diagnoses on subsequent problem solving, Phys. Teach. 54, 87 (2016).
- [34] C. Singh, Categorization of problems to assess and improve proficiency as teacher and learner, Am. J. Phys. 77, 73 (2009).
- [35] A. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, Phys. Rev. ST Phys. Educ. Res. 7, 020110 (2011).
- [36] A. Maries, S. Y. Lin, and C. Singh, Challenges in designing appropriate scaffolding to improve students' representational consistency: The case of a Gauss's law problem, Phys. Rev. Phys. Educ. Res. 13, 020103 (2017).

- [37] C. Singh and D. Haileselassie, Developing problemsolving skills of students taking introductory physics via web-based tutorials, J. Coll. Sci. Teach. **39**, 42 (2010).
- [38] 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 and Science Survey, Phys. Rev. ST Phys. Educ. Res. 2, 010101 (2006).
- [39] C. Singh, What can we learn from PER: Physics Education Research?, Phys. Teach. 52, 568 (2014).
- [40] A. Mason and C. Singh, Using categorization of problems as an instructional tool to help introductory students learn physics, Phys. Educ. 51, 025009 (2016).
- [41] A. Maries and C. Singh, Do students benefit from drawing productive diagrams themselves while solving introductory physics problems? The case of two electrostatics problems, Eur. J. Phys. **39**, 015703 (2018).
- [42] A. Maries and C. Singh, Case of two electrostatics problems: Can providing a diagram adversely impact introductory physics students' problem solving performance?, Phys. Rev. Phys. Educ. Res. 14, 010114 (2018).
- [43] M. Schommer, Effects of beliefs about the nature of knowledge on comprehension, J. Educ. Psychol. 82, 498 (1990).
- [44] K. Cummings, S. Lockwood, and J. Marx, Attitudes toward problem solving as predictors of student success, AIP Conf. Proc. 720, 133 (2004).
- [45] E. Elby, Helping physics student learn how to learn, Am. J. Phys. 69, S54 (2001).
- [46] E. Redish, J. Saul, and R. Steinberg, Student expectations in introductory physics, Am. J. Phys. 66, 212 (1998).
- [47] K. Gray, W. Adams, C. Wieman, and K. Perkins, Students know what physicists believe, but they don't agree: A study using CLASS survey, Phys. Rev. ST Phys. Educ. Res. 4, 020106 (2008).
- [48] J. Marx and K. Cummings, What factors really influence shifts in students' attitudes and expectations in an introductory physics course?, AIP Conf. Proc. 883, 101 (2007).
- [49] A. Mason and C. Singh, Surveying graduate students' attitudes and approaches to problem solving, Phys. Rev. ST Phys. Educ. Res. 6, 020124 (2010).
- [50] A. Mason and C. Singh, Surveying college introductory students' attitudes and approaches to problem solving, Eur. J. Phys. 37, 055704 (2016).
- [51] N. Balta, A. Mason, and C. Singh, Surveying Turkish high school and university students' attitudes and approaches to physics problem solving, Phys. Rev. Phys. Educ. Res. 12, 010129 (2016).
- [52] L. Reddy, An insight into the students' attitude and approach to problem solving in physics at a South African University, Proceedings of the 2016 ISTE International Conference on Mathematics, Science and Technology Education, http://uir.unisa.ac.za/handle/10500/22877.
- [53] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, Phys. Teach. 30, 141 (1992).
- [54] I. A. Hallouin and D. Hestenes, Common sense concepts about motion, Am. J. Phys. 53, 1043 (1985).

- [55] A. Savinainen and P. Scott, The Force Concept Inventory: A tool for monitoring student learning, Phys. Educ. 37, 45 (2002).
- [56] V. Coletta and J. Phillip, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, Am. J. Phys. 73, 1172 (2005).
- [57] M. Lorenzo, C. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. 74, 118 (2006).
- [58] D. Maloney, T. O'Kuma, C. Hieggelke, and A. Van Heuvelen, Surveying students conceptual knowledge of electricity and magnetism, Am. J. Phys. 69, S12 (2001).
- [59] M. Dancy and C. Henderson, Pedagogical practices and instructional change of physics faculty, Am. J. Phys. 78, 1056 (2010).
- [60] C. Henderson and M. Dancy, Impact of physics education research on the teaching of introductory quantitative physics in the United States, Phys. Rev. ST Phys. Educ. Res. 5, 020107 (2009).
- [61] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj, Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovationdecision process?, Phys. Rev. ST Phys. Educ. Res. 8, 020104 (2012).
- [62] R. Hake, Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. 66, 64 (1998); R. R. Hake, Analyzing change/gain scores, https://www .slideshare.net/arvindalalang1/analyzing-change-gainhake (1999).
- [63] E. Mazur, *Peer Instruction: A Users Manual* (Prentice Hall, Upper Saddle River, NJ, 1997); C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).
- [64] A. Mason and C. Singh, Helping students learn effective problem solving strategies by reflecting with peers, Am. J. Phys. 78, 748 (2010).
- [65] A. Mason and C. Singh, Impact of guided reflection with peers on the development of effective problem solving strategies and physics learning, Phys. Teach. 54, 295 (2016).
- [66] D. Sokoloff and R. Thornton, Using interactive lecture demonstrations to create an active learning environment, Phys. Teach. 35, 340 (1997).
- [67] S. Zappe, R. Leicht, J. Messner, T. Litzinger, and H. Lee, Flipping the classroom to explore active learning in a large undergraduate course, *Proceedings of the American Society for Engineering Education Annual Conference & Exposition* (2009).
- [68] N. Karim, A. Maries, and C. Singh, Do evidence-based active-engagement courses reduce the gender gap in introductory physics?, Eur. J. Phys. 39, 025701 (2018).
- [69] N. Karim, A. Maries, and C. Singh, Impact of evidencebased flipped or active-engagement non-flipped courses on student performance in introductory physics, Can. J. Phys. 96, 411 (2018).
- [70] A. Traxler, R. Henderson, J. Stewart, G. Stewart, A. Papak, and R. Lindell, Gender fairness within the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. 14, 010103 (2018).
- [71] A. Traxler, X. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a

complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).

- [72] A. Madsen, S. McKagan, and E. 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).
- [73] K. Kreutzer and A. Boudreaux, Preliminary investigation of instructor effects on gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. 8, 010120 (2012).
- [74] B. Adegoke, Impact of interactive engagement on reducing the gender gap in quantum physics learning outcomes among senior secondary school students, Phys. Educ. 47, 462 (2012).
- [75] A. J. Cavanagh, O. R. Aragón, X. Chen, B. A. Couch, M. F. Durham, A. Bobrownicki, D. I. Hanauer, M. J. Graham, and M. Lee Ledbetter Student buy-in to active learning in a college science course, CBE Life Sci. Educ. 15, ar76 (2016).
- [76] L. Bian, S. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence childrens interests, Science 355, 389 (2017).
- [77] G. Marchand and G. Taasoobshirazi, Stereotype threat and women's performance in physics, Int. J. Sci. Educ. 35, 3050 (2013).
- [78] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. 84, 191 (1977).
- [79] A. Collins, J. Brown, and S. Newman, In Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser, edited by L. B. Resnick (Lawrence Erlbaum, Hillsdale, NJ, 1989), p. 453.
- [80] J. Sweller, Cognitive load during problem solving: Effects on learning, Cogn. Sci. 12, 257 (1988).
- [81] A. Madsen, S. McKagan, and E. Sayre, How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. 11, 010115 (2015).
- [82] T. Nokes-Malach, E. Marshman, Z. Y. Kalender, C. Schunn, and C. Singh, Investigation of male and female students' motivational characteristics throughout an introductory physics course sequence, in *Proceedings of the 2017 Physics Education Research Conference*, Cincinnati, OH, edited by L. Ding, A. Traxler, and Y. Cao (2018), pp. 276–279, https://doi.org/10.1119/perc .2017.pr.064.
- [83] E. Marshman, Z. Y. Kalender, C. Schunn, T. Nokes-Malach, and C. Singh, A longitudinal analysis of underrepresented students' motivational characteristics in introductory physics courses, Can. J. Phys. 96, 391 (2018).
- [84] E. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, Phys. Rev. Phys. Educ. Res. 14, 020123 (2018).
- [85] T. Nokes-Malach, Z. Y. Kalender, E. Marshman, C. Schunn, and C. Singh, Prior preparation and motivational characteristics mediate relations between gender and learning outcomes in introductory physics, in

Proceedings of the 2018 Physics Education Research Conference, Washington, DC, edited by A. Traxler, Y. Cao, and S. Wolf (2018), http://dx.doi.org/10.1119/perc .2018.pr.Nokes-Malach.

- [86] Z. Y. Kalender, E. Marshman, C. Schunn, T. Nokes-Malach, and C. Singh, Large gender differences in physics self-efficacy at equal performance levels: A warning sign?, in *Proceedings of the 2018 Physics Education Research Conference*, Washington, DC, edited by A. Traxler, Y. Cao, and S. Wolf (2018), http://dx.doi .org/10.1119/perc.2018.pr.Kalender.
- [87] Z. Y. Kalender, E. Marshman, T. Nokes-Malach, C. Schunn, and C. Singh, Beliefs About Competence: The Story of Self-Efficacy, Gender, and Physics, in *Diversity Across Disciplines: Research on People, Policy, Process and Paradigm*, eds. A. Murrell and J. Petrie, (Information Age Publishers, Charlotte, NC, 2019).
- [88] A. Maries, N. Karim, and C. Singh, Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?, Phys. Rev. Phys. Educ. Res. 14, 020119 (2018).
- [89] A. Maries, N. Karim, and C. Singh, Does stereotype threat affect female students performance in introductory physics? in AIP Conf. Proc. 2109, 120001 (2019).
- [90] A. Maries, N. Karim, and C. Singh, The impact of stereotype threat on gender gap in introductory physics, in *Proceedings of the 2017 Physics Education Research Conference*, Cincinnati, OH, edited by L. Ding, A. Traxler, and Y. Cao (2018), pp. 256–259, http://dx.doi .org/10.1119/perc.2017.pr.059.
- [91] Z. Y. Kalender, E.Marshman, T. Nokes-Malach, C. Schunn, and C. Singh, Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses, in *Proceedings of the 2017 Physics Education Research Conference, Cincinnati, OH*, edited by L. Ding, A. Traxler, and Y. Cao (2018), pp. 204–207, http://dx.doi.org/10.1119/perc.2017.pr.046.
- [92] J. Cohen, Statistical Power Analysis for the Behavioral Sciences (Routledge, London, 1988).
- [93] R. G. Lomax, Statistical Concepts: A Second Course, 3rd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ, 2007).
- [94] J. Docktor and K. Heller, Gender differences in both Force Concept Inventory and Introductory Physics Performance, AIP Conf. Proc. 1064, 15 (2008).
- [95] C. Steele, A threat in the air: How stereotypes shape the intellectual identities and performance of women and African-Americans, Am. Psychol. 52, 613 (1997).
- [96] S. Beilock, W. Jellison, R. Rydell, A. McConnell, and T. Carr, On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? Personality Soc. Psych. Bull. 38, 1059 (2006).
- [97] S. Beilock, R. Rydell, and A. McConnell, Stereotype threat and working memory: Mechanisms, alleviations, and spillover, J. Exp. Psychol. Gen. 136, 256 (2007).
- [98] D. Yeager and G. Walton, Social-psychological interventions in education: They're not magic, Rev. Educ. Res. 81, 267 (2011).

- [99] C. Dweck, Is Math a Gift? Beliefs that Put Females at Risk, in *Why Arent More Women in Science? Top Researchers Debate the Evidence*, Washington DC, 2006, edited by S. Ceci and W. Williams (American Psychological. Assoc., Washington DC, 2006).
- [100] C. Dweck, *Mindset: The New Psychology of Success*, (Random House Inc., New York, 2006).
- [101] M. Perry, Distributed Cognition in HCI Models, Theories, and Frameworks: Toward an Interdisciplinary Science, edited by J. Carroll (Morgan Kaufmann, 2003), pp. 193–223.