

Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum

Michael J. Cahill,¹ K. Mairin Hynes,² Rebecca Trousil,³ Lisa A. Brooks,⁴ Mark A. McDaniel,^{1,5}
Michelle Repice,^{1,6} Jiuqing Zhao,¹ and Regina F. Frey^{1,6,7,*}

¹*Center for Integrative Research on Cognition, Learning, and Education,
Washington University in St. Louis, One Brookings Drive, Saint Louis, Missouri 63130, USA*

²*Department of Physics, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

³*Department of Mathematics and Computer Science, Skidmore College,
815 North Broadway, Saratoga Springs, New York 12866, USA*

⁴*Acumen Research and Evaluation, LLC, 1811 North Reynolds Road, Toledo, Ohio 43615, USA*

⁵*Department of Psychology, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

⁶*The Teaching Center, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

⁷*Department of Chemistry, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

(Received 19 February 2014; published 2 July 2014)

Interactive-engagement (IE) techniques consistently enhance conceptual learning gains relative to traditional-lecture courses, but attitudinal gains typically emerge only in small, inquiry-based curricula. The current study evaluated whether a “scalable IE” curriculum—a curriculum used in a large course (~130 students per section) and likely adoptable by a wide range of physics departments—could produce significant attitudinal benefits relative to a traditional-lecture curriculum. This study included data across three years, 10 instructors, over 30 sections, and over 1100 students, and our analytic strategy allowed us to isolate the effects that were due to the curriculum itself rather than other potential factors such as instructor differences or preexisting differences among students. Results revealed that our Active-Physics curriculum, which is based on Moore’s *Six Ideas That Shaped Physics*, produced significant attitudinal and conceptual-learning benefits relative to our traditional-lecture physics curriculum. Further, the Active-Physics curriculum, for the most part, benefitted males and females equally, and relative to the Fall semester alone, the benefits of Active Physics became more robust when viewed across the entire two-semester sequence of introductory physics. Our data highlight that some (though not all) of the attitudinal benefits of small, inquiry-based courses may be achievable in larger course with scalable IE curricula that can potentially reach a large proportion of introductory physics students.

DOI: [10.1103/PhysRevSTPER.10.020101](https://doi.org/10.1103/PhysRevSTPER.10.020101)

PACS numbers: 01.40.G-

I. INTRODUCTION

A sizable body of literature has emerged examining the impact of various curricula on introductory physics students’ conceptual understanding [1,2] and attitudes (typically defined in this literature as the extent to which students endorse expertlike perceptions of and approaches toward physics) [3,4]. A major focus of this literature has been the effects of interactive-engagement (IE) techniques on student outcomes. Hake [1] defines IE techniques as those “designed at least in part to promote conceptual

understanding through ... heads-on (always) and hands-on (usually) activities which yield immediate feedback through discussion with peers and/or instructors” (p. 65), and previous studies have reported robust benefits of IE on conceptual learning [1,2]. However, attitudinal gains have been elusive, even in IE courses, and only recently have reports of sizable attitudinal gains emerged [5–8]. These gains have come primarily from “inquiry-based” curricula—those that completely replace lecture with IE activities (typically small-group, inquiry-based activities).

Although the benefits of inquiry-based curricula such as Modeling Instruction [5,6], Physics and Everyday Thinking/Physical Science and Everyday Thinking [7], and Physics by Inquiry [8] are clear, there are practical constraints that severely limit the proportion of students with access to these curricula. For example, inquiry-based curricula may not be feasible or effective in large classes. Also, some instructors may be resistant to eliminating

*Corresponding author.
gfrey@wustl.edu

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

lecture completely in order to adopt pure inquiry-based curricula. Consequently, the majority of students are unlikely to experience an inquiry-based curriculum and the benefits therein. For example, at Florida International University only 25% of introductory physics students who attempt to register for Modeling Instruction are allowed to take the course, with a lottery system implemented to select these students [6]. Thus, a crucial objective should be to develop and optimize “scalable” curricula—those that possess an overall structure that makes them tractable in large classes and more likely to achieve widespread adoption.

Recent reforms, including SCALE-UP [9] and research-based transformations at the University of Colorado [3,10] and Harvard [11], have aimed to integrate effective IE activities into large classrooms, and this work has shown promise in conferring the benefits of IE in large courses. The current study aims to show further evidence of the merits of these curricula with a systematic, large-scale, multiyear investigation of the effects of one such scalable IE curriculum—the Active-Physics curriculum at Washington University in St. Louis (which is based on Moore’s *Six Ideas That Shaped Physics* [12]). We directly compared the outcomes of students in Active Physics with those in traditional-lecture physics during the same three-year span, and our investigation was organized around the following four research questions:

- (1) Does Active Physics improve conceptual learning relative to traditional-lecture physics?
- (2) Does Active Physics improve attitudes relative to traditional-lecture physics?
- (3) Does the impact of Active Physics vary across gender?
- (4) Relative to traditional-lecture physics, how does Active Physics impact attitudes across the whole year?

A. Does Active Physics improve conceptual learning, relative to traditional-lecture physics?

In a review spanning 62 introductory physics courses across a range of high schools, colleges, and universities, Hake [1] demonstrated that courses including IE produced approximately double the gains of traditional-lecture courses on conceptual understanding of mechanics as measured by either the Mechanics Diagnostic (MD) [13] or Force Concept Inventory (FCI) [14]. In Hake’s study the benefits of IE on conceptual learning were robust across a variety of conceptual learning instruments, curricula, populations, and class sizes. Additionally, more recent investigations have shown the same pattern when comparing large traditional-lecture courses to scalable IE courses [2,3]. These effects have been found both across an entire semester [3] and within a single class period [2]. Based on this previous literature, we hypothesized that Active Physics would produce an advantage in conceptual learning gains relative to traditional-lecture physics.

B. Does Active Physics improve attitudes relative to traditional-lecture physics?

Previous research has reported steep attitudinal declines in traditional-lecture introductory physics courses [3,15] motivating reforms aimed at reversing or at least stopping these declines. Recently, impressive gains on the Colorado Learning Attitudes about Science Survey (CLASS) [16] have emerged in pure inquiry-based introductory physics courses. Overall, CLASS gains ranging from 3% to 12% have been achieved with the Modeling Instruction curriculum [5,6,17], 4% to 16% with Physics in Everyday Thinking/Physical Science in Everyday Thinking curricula [7], and 1% to 25% with Physics by Inquiry [8].

Inquiry-based courses clearly benefit attitudes but may not be scalable to large classes and may be unlikely to achieve widespread adoption. First, at many universities introductory physics courses include hundreds of students, and the practical feasibility and effectiveness of inquiry-based curricula have not been demonstrated for classes of those sizes [6]. Enrollment for inquiry-based courses reported in the literature is typically below 30 and almost always below 50, with only one larger inquiry-based course (100 students) included in the literature [7]. Second, these courses may not allow for the breadth of coverage that traditional courses allow. Some inquiry-based curricula, such as Physics in Everyday Thinking and Physics by Inquiry, were designed for nonscience majors and thus the ability to cover a large amount of physics content was not a primary concern. However, for physics and other science majors, covering all of the material that will be built upon in future courses in the sequence may be a larger concern. Also, instructors may be resistant to the idea of such a drastic overhaul from traditional-lecture to inquiry-based curricula. In examining instructor resistance to research-based instructional techniques, the primary reasons for resistance included the time necessary to learn or implement new strategies, pressure to cover a large amount of content, and perceived student resistance, among others [18,19]. Time and the pressure to cover a large amount of content may be particularly strong barriers in the context of inquiry-based curricula. Because of these practical impediments, inquiry-based curricula may not soon be available to a sizable portion of introductory physics students. In contrast, Active Physics can be implemented in large classes (up to 130 so far), can be supplemented with ready-made activities to reduce time strain, and covers the same breadth of content as traditional-lecture physics (the issue of student resistance will be revisited in Sec. IV). Thus, Active Physics (and other scalable IE curricula) minimizes the implementation barriers present in inquiry-based curricula and may potentially achieve relatively widespread adoption.

With these practical issues in mind, one of our primary questions was whether at least *some* of the attitudinal gains associated with inquiry-based curricula could be achieved with a scalable IE curriculum such as Active

Physics. Previous reports with scalable IE curricula present encouraging but mixed results [3,4,10,20]. An initial demonstration using the CLASS reported slight (1.5%) increases across the semester in interactive-engagement courses, compared to sharp declines (−9.8%) for concurrent traditional-lecture courses [3]. However, Pollock [4] demonstrated that the effectiveness of the interactive-engagement curriculum varied with slight changes in implementation, and a broader look at the effect of similar classes reveals a mixed pattern of results, with instances of both CLASS gains [3,4,20] and CLASS decreases [4,10,20].

One multisection demonstration of large attitudinal increases in large classes (100–200 students) [21] involved a curriculum that, despite its effectiveness in large classes, may not be “scalable” for other reasons. Implementing this curriculum involves changes beyond the main section, including careful alteration of recitation and lab sections and the addition of a teaching assistant (TA)-staffed “course center” where students can gather to complete homework. Such a comprehensive transformation requires immense departmental investment and may not be feasible within physics departments at most universities. In contrast, implementing Active Physics involves altering only the main section and does not require extra staffing—both Active-Physics and traditional-lecture physics sections are staffed by one instructor and one teaching assistant.

SCALE-UP is another reform shown to produce attitudinal gains (in addition to impressive conceptual learning gains) in large introductory physics classes [9]. Although this reform is designed specifically for large classes and has been used with great effect at many universities, it requires a complete redesign of the classroom and significant technological resources. Many universities may not have the resources to implement SCALE-UP. Thus, it is important to determine whether less resource-dependent curricula can produce some (though perhaps not all) of the same benefits in large classes.

The current study examined whether a scalable IE curriculum—Active Physics (based on Moore’s *Six Ideas That Shaped Physics* [12])—could benefit students’ attitudes. We examined absolute CLASS shifts, but in light of previous research, we made no strong predictions about what form this absolute shift would take. The more important issue from our perspective was how these shifts compared to the traditional-lecture curriculum. We predicted that, regardless of whether Active Physics produced positive, negative, or neutral CLASS shifts in absolute terms, this curriculum would produce a benefit on CLASS attitudes relative to the effect in the traditional-lecture curriculum.

C. Does the impact of Active Physics vary across gender?

Males tend to begin introductory physics with more conceptual knowledge than females, as measured by the Force Concept Inventory (FCI) [11] and Force and Motion

Concept Evaluation (FMCE) [10,22], but the current literature provides somewhat mixed results regarding the influence of scalable IE courses on this gap [23]. In this literature a distinction is made between courses in which interactive engagement (IE) activities are part of the lecture only (IE 1) or part of both the lecture and subsections (IE 2). Initial research demonstrated that, by semester’s end, the preexisting gender gap in FCI was slightly reduced in IE 1 courses and eliminated in IE 2 courses (despite the fact that both genders benefitted) [11]. However, later findings raised concerns that these results do not generalize to all populations implementing similar curricula and reported that the gender gap on the FMCE decreased in IE 2 courses but actually increased in IE 1 courses [10,22]. Further, these later studies examined normalized gains—the gain made relative to the maximum gain possible—which is usually the measure of interest for concept inventories in literature [1] and reported that these normalized gains were actually higher for males in both IE 1 and IE 2 courses [10,22]. In addition, males enter physics with a more positive CLASS score than females, and this gap increases in both IE 1 and IE 2 courses [10].

Overall, the influence of IE on gender gaps is highly variable with gender gaps sometimes increasing across scalable IE introductory physics courses [23]. Findings such as these might cause hesitation in adopting IE elements into one’s course. However, it is important to note that, even in cases where gender gaps have increased in scalable IE courses, there is no evidence that the IE elements *caused* gender gaps to increase or in any way impacted females negatively. In fact, in these cases it is entirely possible that scalable IE courses boosted female performance and reduced the gender gap relative to what would have occurred in a traditional-lecture course. In the current study we had a concurrent sample of introductory physics students taught in a traditional-lecture course. This allowed us to examine the impact of our scalable IE curriculum on females’ performance and gender gaps relative to this traditional-lecture course. We made no prediction regarding the absolute changes in attitudinal and conceptual learning gender gaps, per se, within Active Physics. However, we predicted that relative to traditional-lecture physics, Active Physics would benefit both males and females on both conceptual learning and attitudes and potentially lead to smaller gender gaps on these outcomes.

D. Relative to traditional-lecture physics, how does Active Physics impact attitudes across the whole year?

The majority of evaluations of interactive-engagement curricula are limited to the Fall semester [3,7,10,20]. The few studies examining the CLASS across the second semester of introductory physics, covering electricity and magnetism, typically have revealed attitudinal decreases even in scalable IE courses [4,24]. A slight CLASS increase (~3%) from pre-Spring to post-Spring has been achieved in

a Modeling Instruction course [5], but this increase was negligible compared to the large CLASS gains ($\sim 9\%$) produced by Modeling Instruction in the Fall semester within the same study. It is perhaps not surprising that students' perceptions become less expertlike during this semester even in scalable IE courses, given the nature of the material covered during the semester. Relative to mechanics, electricity and magnetism may be difficult to grasp conceptually during a single semester, and students thus may take more algorithmic approaches to learning the material and adopt less expertlike perceptions of physics. However, despite absolute decreases across the semester, scalable IE courses potentially could be providing some benefit relative to what might occur in a traditional-lecture course. The present study allowed for direct comparison between the Active-Physics curriculum and a concurrent traditional-lecture curriculum. Based on previous literature, we suspected that the CLASS would, in absolute terms, decrease across the Spring semester, but we predicted that this decrease would be significantly less pronounced in the Active-Physics class.

II. METHOD

A. Methodological notes

The present study contained a number of methodological characteristics that allowed us to assess the effects of the Active-Physics curriculum in a rigorous manner. First, due to the fortunate timing of our investigation, the current study included a good “baseline” comparison group—a concurrent traditional-lecture introductory physics curriculum drawn from the same population as our Active-Physics sample and possessing similar demographic profiles. Direct statistical comparisons were made between the two curricula, providing a reference point from which to evaluate the learning gains and attitudinal shifts that occurred in Active Physics. Second, this was a relatively large-scale study, with large N (1120 Fall-only and 921 whole-year participants), a large number of sections (22 Active and 10 traditional lecture across Fall and Spring), and a variety of instructors (5 Active, 5 traditional lecture, and 2 who taught both). Implementation and effectiveness can vary across sections and instructors [4], but the results in aggregate should accurately represent the impact that these two curricula have on learning and attitudes. Finally, we utilized statistical techniques (described in detail below) to control for instructor effects and other variables besides curriculum that could have contributed to differences between the traditional-lecture and Active-Physics samples (i.e., confounds). Examples of such confounds controlled in the present study include individual and classwide differences on our pretest measures. These methodological steps allowed us to better isolate the true effects of the Active-Physics curriculum relative to the traditional-lecture baseline curriculum.

B. Sample

Our study included students who took the two-semester introductory physics sequence at Washington University in St. Louis during the 2009–2010, 2010–2011, and 2011–2012 academic years. Across these three years there were 12 total instructors (5 Active Physics, 5 traditional-lecture physics, and 2 who taught both) and a total of 2035 students (1236 in Active Physics, 796 in traditional-lecture physics, and 3 who took Active Physics in the Fall and traditional-lecture physics in the Spring). (Students were not allowed to migrate from traditional-lecture physics in the Fall to Active Physics in the Spring, and the three students who migrated from Active Physics in the Fall to traditional lecture in the Spring were not included in the whole-year analyses.) Across the three years, both the Fall and Spring semesters included 11 Active-Physics and 5 traditional-lecture sections. Students in the two curricula were drawn from the same population of students, as both curricula were available to all students who were part of the study. However, students self-selected into the curricula, introducing potential nonrandom differences between the two groups of students.

Student demographics are presented in Tables I and II. Because some students participated in only certain parts of the study, the number of students (N) and consequently the demographic breakdown differ for different analyses. We have provided the demographics for the following five overlapping subsets of the sample: (1) pre-Fall: students who took the CLASS [16] at the beginning of the Fall semester; (2) Fall only: students who took the CLASS at both the beginning and end of the Fall semester; (3) whole year: students who took the CLASS at the beginning of Fall and at the end of Spring; (4) FCI: students who took the FCI [14] at the beginning and end of the Fall semester; (5) BEMA: students who took the Brief Electricity and Magnetism Assessment (BEMA) [25] at the beginning and end of the Spring semester.

For each of the groups listed above, percentages for the descriptive characteristics of the students such as gender, number of underrepresentative minorities, majors, and ACT scores are provided in Tables I and II; these demographics were obtained from the university student-information database. Table I contains descriptive characteristics of students who completed the CLASS; Table II compares characteristics of students who completed the FCI or the BEMA.

For all five groups we compared the demographics of the students between Active Physics and traditional-lecture physics. For all five samples there were slightly greater percentages of females in the Active-Physics sections, but the difference was not statistically significant in any of the samples (χ^2 range from 0.01 to 1.06, p 's range from 0.30 to 0.94). All five samples also showed slightly lower percentages of minorities in the Active-Physics sections, but the difference was not significant (χ^2 range from 0.24 to

TABLE I. Demographics for students taking the CLASS survey. Bold cells indicate a significant difference between traditional-lecture physics and Active Physics. SE indicates Standard error.

	Pre-Fall sample			Fall-only sample			Whole-year sample		
	Lecture	Active	p^a	Lecture	Active	p	Lecture	Active	p
Number of students in sample	451	929	<0.001	357	763	<0.001	250	671	0.02
% of students in class	66.3	76.0		52.5	62.4		54.3	60.7	
Student characteristics									
	Lecture	Active	p^a	Lecture	Active	p	Lecture	Active	p
% Female	39.2	41.0	0.53	40.3	42.2	0.56	38.0	41.0	0.41
% Under-represented minorities ^b	11.2	10.2	.57	11.4	10.0	0.50	12.8	10.4	0.32
% 1st or 2nd year students	15.3	30.5	<0.001	16.0	30.9	<0.001	16.4	32.0	<0.001
Majors ^c									
% Arts and sciences STEM ^d	20.8	37.4	<0.001	21.8	37.7	<0.001	24.4	39.6	<0.001
% Arts & sciences non-STEM	23.9	21.4		21.8	22.8		19.6	21.5	
% Engineering	41.5	38.4		42.6	37.1		53.2	37.6	
% Other	13.7	2.8		13.7	2.4		2.8	1.3	
College entrance exam ACT scores ^e									
	Lecture ($n = 447$)		Active ($n = 928$)	Lecture ($n = 354$)		Active ($n = 763$)	Lecture ($n = 249$)		Active ($n = 671$)
	Mean (SE)	Mean (SE)	p^f	Mean (SE)	Mean (SE)	p	Mean (SE)	Mean (SE)	p
ACT verbal ^e	32.32 (0.11)	32.87 (0.07)	<0.001	32.51 (0.12)	32.91 (0.07)	0.003	32.50 (0.14)	32.92 (0.08)	0.008
ACT math ^g	33.17 (0.11)	33.57 (0.07)	0.001	33.23 (0.12)	33.55 (0.08)	0.02	33.39 (0.13)	33.55 (0.08)	0.31

^a p values for % female, % underrepresented minority, and year in school are derived from χ^2 tests for independence conducted with instruction type as one categorical variable and the particular demographic variable (gender, race, and year in school) as the second categorical variable. A significant p value indicates that traditional-lecture physics and Active Physics differ in the proportion of a given category.

^bMinority students include the following groups: Black, Non-Hispanic; American Indian or Native Alaskan; Hispanic; and Multiracial minorities. International students were not considered to be underrepresented minorities. Percentages are based on students whose race was reported. The number of students not reporting race varies as follows: 33 traditional-lecture physics and 47 Active Physics in pre-Fall sample; 23 traditional-lecture physics and 35 Active Physics in Fall-only sample; 15 traditional-lecture physics and 36 Active Physics in whole-year sample.

^cMajors were subject to $2(\text{instruction type}) \times 4(\text{majors})$ χ^2 tests for independence. A significant p value indicates that majors are distributed differently across Active Physics and traditional-lecture physics. The test does not specify which proportions significantly differ from one another—only that the pattern of proportions does differ across Active Physics and traditional-lecture physics.

^dArts and sciences STEM includes biology, chemistry, Earth and planetary sciences, environmental biology, environmental Earth sciences, mathematics, and physics.

^eACT verbal scores equal the average of ACT English and ACT reading. For students without these ACT scores, concordance tables were used to convert SAT verbal to the sum of ACT English and ACT reading [26]. The resulting sum was then divided by 2 to put it on a 0–36 scale.

^f p values for ACT verbal and ACT math derived from independent samples t -tests with instruction type as the independent variable.

^gFor students without ACT math scores, concordance tables were used to convert SAT math scores to ACT math scores [26].

1.86, p 's range from 0.17 to 0.62). The proportion of 1st and 2nd year students was significantly higher in Active-Physics sections across all five samples (χ^2 range from 22.20 to 36.74, all $p < 0.001$). In all five samples the distributions of majors were statistically significantly different between Active Physics and traditional-lecture physics (χ^2 range from 25.11 to 84.55, all $p < 0.001$). These differences are driven by a higher proportion of arts and

sciences Science-Technology-Engineering-Mathematics (STEM) students in Active Physics, and a higher proportion of engineers and “other” students (mostly from business and architecture) in traditional-lecture physics.

In all five samples Active Physics held a statistically significant advantage in ACT verbal (t ranges from 2.01 to 3.42, p ranges from <0.001 to 0.045). However, as seen in Tables I and II, in traditional-lecture physics the means

TABLE II. Demographics for students taking the FCI and BEMA. Bold cells indicate a significant difference between traditional-lecture physics and Active Physics. SE indicates Standard error.

	FCI sample			BEMA Sample				
	Lecture	Active	p^a	Lecture	Active	p		
Number of students in sample	366	773	<0.001	360	738	0.14		
% of students in class	53.8	63.3		62.2	65.8			
Student characteristics								
	Lecture	Active	p^a	Lecture	Active	p		
% Female	39.3	42.6	0.30	40.6	40.8	0.94		
% URM ^b	11.0	9.7	0.50	11.1	10.6	0.79		
% 1st or 2nd year	15.3	29.9	<0.001	19.2	33.2	<0.001		
	Major ^c							
% Arts and sciences STEM ^d	23.0	37.3	<0.001	27.8	39.8	<0.001		
% Arts and sciences non-STEM	21.0	22.6		17.2	21.4			
% Engineering	43.2	37.8		51.7	37.3			
% Other	12.8	2.3		3.3	1.5			
College entrance exam ACT scores ^e								
	Lecture ($n = 362$)		Active ($n = 773$)		Lecture ($n = 358$)		Active ($n = 738$)	
	Mean (SE)	Mean (SE)	p^f	Mean (SE)	Mean (SE)	p		
ACT verbal ^e	32.51 (0.12)	32.91 (0.07)	0.004	32.49 (0.12)	32.86 (0.08)	0.007		
ACT math ^g	33.27 (0.11)	33.58 (0.08)	0.03	33.50 (0.11)	33.49 (0.08)	0.97		

^a p values for % female, % underrepresented minority, and year in school are derived from χ^2 tests for independence conducted with instruction type as one categorical variable and the particular demographic variable (gender, race, and year in school) as the second categorical variable. A significant p value indicates that traditional-lecture physics and Active Physics differ in the proportion of a given category.

^bMinority students include the following groups: Black, Non-Hispanic; American Indian or Native Alaskan; Hispanic; and Multiracial minorities. International students were not considered to be underrepresented minorities. Percentages are based on students whose race was reported. The number of students not reporting race varies as follows: 21 traditional-lecture physics and 41 Active Physics in FCI sample; 19 traditional-lecture physics and 41 Active Physics in BEMA sample.

^cMajors were subject to $2(\text{instruction type}) \times 4(\text{majors}) \chi^2$ tests for independence. A significant p value indicates that majors are distributed differently across Active Physics and traditional-lecture physics. The test does not specify which proportions significantly differ from one another—only that the pattern of proportions does differ across Active Physics and traditional-lecture physics.

^dArts and sciences STEM includes biology, chemistry, Earth and planetary sciences, environmental biology, environmental Earth sciences, mathematics, and physics.

^eACT verbal scores equal the average of ACT English and ACT reading. For students without these ACT scores, concordance tables were used to convert SAT verbal to the sum of ACT English and ACT reading [26]. The resulting sum was then divided by 2 to put it on a 0–36 scale.

^f p values for ACT verbal and ACT math derived from independent samples t -tests with instruction type as the independent variable.

^gFor students without ACT math scores, concordance tables were used to convert SAT math scores to ACT math scores [26].

range from 31.83 to 32.08, and in Active Physics the means range from 32.40 to 32.47; these differences are unlikely to have any practical significance. Active Physics showed a statistically significant advantage on ACT math for the pre-Fall, Fall-only, and FCI samples. In all cases the absolute differences in ACT math scores were minimal; for both traditional-lecture physics and Active Physics, the means were between 33.17 and 33.58. Further, there were no significant differences in ACT math for the whole year ($t = 1.02$, $p = 0.31$) or BEMA ($t = 0.04$, $p = 0.97$). Thus, differences between math ACT are unlikely to be of any practical significance. Hence, overall, the Active-Physics and traditional-lecture samples had somewhat

different distributions of majors but were very similar on all of the other demographic measures.

C. Curricula

1. Traditional-lecture curriculum

The textbook used in traditional-lecture physics was *Sears and Zemansky's University Physics with Modern Physics* [27]. The Fall semester included 27-class hours (approximately 75% out of 36.5-class hours) devoted to mechanics, with the remainder spent covering relativity, acoustics, and fluids. In the Spring semester, 27-class hours (approximately 65% out of 41.5-class hours) were spent on electromagnetism, with the remainder covering an introduction to quantum

mechanics and nuclear physics. During each semester three noncumulative exams were given.

Traditional-lecture physics was held in a lecture hall, and enrollment ranged from 100 to 180 (typical enrollment ~ 140). Class time was spent primarily covering material in lecture style (including derivations and problem solving on the board) with intermittent demonstrations. Prior to each lecture students were assigned reading from the textbook, but no homework was assigned prior to the lecture. Weekly graded homework assignments were due approximately one week after the material was covered in class. Additional practice problems were suggested but not required. The weekly homework and additional practice problems were end-of-chapter textbook problems; no nontextbook problems were given.

2. Active-Physics curriculum

The Active-Physics textbook was *Six Ideas That Shaped Physics* [12]. As with traditional-lecture physics, 27-class hours of the Fall semester (75% out of 36-class hours) were devoted to mechanics with the remainder spent on relativity. In the Spring semester, 17 class hours (approximately 45% out of 38.5-class hours) were spent on electromagnetism, with the remainder of class periods spent on quantum mechanics, nuclear physics, and statistical mechanics. During each semester three noncumulative exams were given, comparable in difficulty to those in traditional-lecture physics.

The classroom was the same lecture hall as traditional-lecture physics. The 180-capacity room allowed the students (typical enrollment 130) room to spread out into groups. Active Physics was grounded in the philosophy that, “Physics is not a collection of facts to absorb, but rather a set of *thinking skills* requiring practice to master” [12]. As such, Active Physics differed from traditional-lecture physics in several ways aimed toward developing deeper conceptual thinking skills as opposed to rote knowledge or purely quantitative problem-solving skills (see Table III for a summary of these differences). The most obvious change is the increase in IE activities during class time. A typical hour-long class includes 2–3 group IE

activities ($\sim 50\%$ of class time). Groups were self-selected and could change from class to class; this informality eased the administrative workload since the instructors did not have to assign groups. Students typically worked in groups of 2–3 but could choose to work alone or in groups of 4. Some of the IE activities were based around demonstrations in class. Traditional-lecture physics and Active Physics had the same demonstrations, but in Active Physics these demonstrations were integrated into a group activity. For example, as an illustration of Archimedes’ principle both curricula included a demonstration in which a small weight displaced more water when placed in a floating “boat” than when placed directly in water. In Active Physics, prior to the demonstration, groups were given time (~ 2 –5 minutes, depending on the demonstration) to discuss and make predictions about the effect of moving the weight from the boat into the water. Individual students (as many as would like to share, typically ~ 3) then volunteered their predictions and reasoning to the rest of the class, and then the demonstration was conducted. No such group or classwide discussion was included in traditional-lecture physics. Other IE activities took the form of 2-minute problems, which are multiple-choice or true-false problems that focus on conceptual ideas and misconceptions in the chapter. Again, students discussed the problems in small groups before volunteering their answers and reasoning to the class as a whole. Instructors then presented and explained the correct answer. (The majority of these problems were taken from the textbook [12] or Mazur’s ConcepTests [28], although some were constructed by the Active-Physics instructors.) Students discussed the problems in small groups for several minutes and then voted on the answer. Groups then volunteered to explain to the class how they thought about and solved the problem, allowing students to learn from their peers’ thought processes.

Lecture was limited to 10–15 minute mini lectures ($\sim 15\%$ of class time). These lectures did not attempt to cover all of the day’s material but rather focused in depth on a particularly challenging or important topic of the day. Little to no class time was spent on derivations. Because only a portion of the course’s content was addressed in class via mini lectures or IE activities, students were responsible

TABLE III. Summary of similarities and differences between Active Physics and traditional-lecture physics.

	Active	Lecture
Text	<i>Six Ideas That Shaped Physics</i>	<i>Sears and Zemansky’s University Physics</i>
Classroom activities	Lecture, problem solving, demo-centered group activities, 2-minute problems	Lecture, derivations, problem solving, and demos (with no group activities)
Problem-solving style	Conceptual and quantitative	Quantitative focus
Daily homework	Yes, on material that was in required reading but not yet covered in class	No
Weekly graded homework	Yes	Yes
Opportunity to revise homework	Yes	No
Exams	3 noncumulative unit exams	3 noncumulative unit exams
Labs	Traditional	Traditional

for learning much of the material outside of class. Homework was used as a tool to ensure that students kept up with the material. Students in both traditional-lecture physics and Active Physics were assigned readings prior to each class, but only students in Active Physics were required to turn in daily homework each class based on this reading. Daily homework problems were usually lower-level (*synthetic*) problems taken from the *Six Ideas* textbook and were selected to ensure that students had a sufficient level of understanding of the necessary information upon which the day's class activities could build a deeper level of conceptual understanding. Weekly homework consisted of higher-level problems (*rich context* for Active Physics) and was comparable for traditional-lecture and Active physics. Another important difference was in grading and feedback. In Active Physics, for both daily and weekly homework, students were assigned points for problems but were not given the correct answer or any specific feedback from the grader. They were, however, given the opportunity to revise their homework and turn it in to potentially gain more points. This encouraged students to seek out the correct answers on their own (answers were provided online) and to figure out why points were lost, providing an additional learning opportunity. It also had the added benefit of making scoring faster for teaching assistants, making the workload comparable to the teaching assistants in traditional-lecture physics despite the addition of daily homework. In Traditional-lecture physics students were given specific feedback on homework and did not have the opportunity to revise.

The remaining portion (~35%) of class time in Active Physics was spent solving problems. Both Active Physics and traditional-lecture physics provided rigorous instruction on quantitative problem-solving skills, but in Active Physics these quantitative skills were embedded within a larger conceptual framework [12]. Problems in Active Physics typically were introduced via real-world questions, and before setting up equations and solving for variables, students first translated the problem into pictures with mathematical symbols and built a conceptual model to determine which physics concepts were relevant. Only then did students perform the quantitative steps of setting up and solving the relevant equations. In traditional-lecture physics students were taught to set up and solve equations for different problem types, but this problem-solving process was not explicitly connected to a more general conceptual framework. In both courses, once a solution was reached, students were asked to determine whether the solution was plausible in the real world.

D. Study instruments

1. Force concept inventory

The FCI [14] is a 30-item multiple-choice test assessing basic knowledge of mechanics concepts covered during the Fall semester. Students completed the FCI at both the beginning and end of the Fall semester. We assessed

changes in the FCI using a measure of *normalized change c* [29]. For students with FCI increasing from pretest to post-test, c is calculated in the following manner:

$$c = \frac{\text{post} - \text{pre}}{100 - \text{pre}}.$$

For students with increasing FCI, c represents the ratio of gains made to the maximum gain that could have been made. This equation is equivalent to the measure of normalized gain g , which is typically used in the literature [1]. The difference between g and c is that g uses the above equation for all students, while c uses the following corrected equation for students who demonstrated decreases from pretest to post-test:

$$c = \frac{\text{post} - \text{pre}}{\text{pre}}.$$

Hence, for students with decreasing FCI, c is the ratio of losses to the maximum loss that could have occurred. To illustrate, a student who decreased from 80% to 40% would have $g = -2$ (a loss of 40% relative to the maximum possible gain of 20%) and $c = -0.5$ (a loss of 40% relative to the maximum possible loss of 80%). The vast majority of students improve from pre- to post-test, so the difference between g and c is usually minimal when averaged across large groups of students, but we are using c because it has a range of -1 to 1 , whereas g can result in large negative values (theoretically ranging to $-\infty$), giving those values undue influence and introducing additional noise in the data.

2. Brief electricity and magnetism assessment

The BEMA [25] is a 30-item multiple-choice test assessing basic knowledge of electromagnetism concepts covered during the second semester of introductory physics. Students completed the BEMA at both the beginning and end of the Spring semester, and as with the FCI, changes in the BEMA were assessed with c .

3. Colorado learning attitudes about science survey

Students completed the CLASS [16] at the beginning of the Fall semester, at the end of the Fall semester, and again at the end of the Spring semester. The CLASS is a 42-item questionnaire assessing students' agreement with statements about their attitudes toward and perceptions about physics and about learning physics. The scale is 5-point Likert scale anchored at 1 (strongly disagree) and 5 (strongly agree).

Of the 42 items, one item (item 31) was used to discard students who are not reading the questions ("We use this statement to discard the survey of people who are not reading the questions. Please select agree-option 4 for this question to preserve your answers."). In addition, five items are not diagnostic due to a lack of consensus among physics experts, and these items were excluded from analysis. The

remaining 36 items had expert consensus, and we used these 36 items in our analysis. Although students responded on a 5-point scale, we coded responses on these items dichotomously as either in agreement with expert consensus (“positive”) or in disagreement (“not positive”) [3,16]. For items with which experts agreed, student responses of 4 or 5 were coded as positive and all other responses as not positive; for items with which experts disagreed, student responses of 1 or 2 were coded as positive and all other responses as not positive. For the overall CLASS survey and each subscale reported, the score represents the percentage of the given items within the scale that were coded as positive, or the *percent positive*. The percent positive of the 36-diagnostic items represents the overall CLASS score.

- (a) *Previously identified subscales*. In addition to the overall score, previous research [16] has identified the following eight subcategories: Real-World Connection, Personal Interest, Sense-Making/Effort, Conceptual Understanding, Applied Conceptual Understanding, Problem-Solving General, Problem-Solving Confidence, and Problem-Solving Sophistication. The score on each subscale represents the percent positive of the responses to the subscale’s items. Factor analyses have confirmed that each subscale in isolation is a statistically robust factor [16]. However, the factors are not independent from each other, and in fact there is a high degree of item overlap across the scales.
- (b) *Newly developed factors*. As a complement to the previous approach [16], we developed a set of orthogonal factors [30] that captured the aspect of the CLASS in which we were most interested: whether students adopted a rote or algorithmic approach or more of a conceptual approach to learning and to problem solving. We selected 25 items that we believed described this aspect. A factor analysis revealed that the 25 items produced two statistically robust factors that are nearly orthogonal. Both factors had 13 items (only item 13 loaded on highly on both factors), with all factor loadings above 0.28 (ranging from 0.28 to 0.66). Appendix A identifies the items in factors 1 and 2, which we have labeled Learning Approach (rote versus conceptual) and Solving Approach (algorithmic versus concept based), respectively. Cronbach’s α for Learning Approach and Solving Approach are 0.81 and 0.76, respectively.

E. Procedure

All introductory physics students were offered a chance in each semester to replace one nonzero lab grade with a perfect score by completing all study instruments within that semester. Students completed all instruments online. In the Fall semester, students completed both the FCI and CLASS during weeks 2 and 3 of the semester and again 2–3 weeks prior to semester’s end. In the Spring semester

students completed the BEMA during weeks 2 and 3 of the semester and completed the BEMA and CLASS 2–3 weeks prior to semester’s end.

F. Analytic strategy

1. Hierarchical regression model

We conducted hierarchical-linear-regression analyses that included two student-level variables (gender and pretest scores on the given dependent variable) and two class-level variables (pretest class mean and instruction type). The instruction-type variable compared Active Physics and traditional-lecture physics. These four variables along with all possible interactions among them were entered as predictors in a regression for each dependent variable (FCI, BEMA, overall CLASS, and all CLASS subscales). This allowed us to control for variables, such as pretest scores of the individual, or preexisting differences between the classes as a whole, that could potentially influence outcome variables and contribute to differences in outcomes of Active-Physics and traditional-lecture students. In all, each model included 15 predictors, consisting of the 4 main effects (gender, individual pretest score, pretest class mean, and instruction type) and all possible interactions among them. All of these terms were included to allow us to isolate the effects of primary interest (instruction-type main effect and instruction-type \times gender interaction) after controlling for all of these other potential sources of variance (all of the other main effects and interactions included in the regression).

The terms central to our study objectives were the instruction-type main effect and the gender \times instruction-type interaction. Instruction-type main effects indicate whether the curricula had a significant overall impact for a given variable (with positive B values indicating an Active-Physics advantage). The instruction-type \times gender interaction terms indicate whether or not the influence of curriculum was consistent across genders. For measures with nonsignificant instruction-type \times gender interaction terms, we concluded that the instruction-type effect was equivalent for males and females. For measures with significant instruction-type \times gender interaction terms, we conducted follow-up analyses within each gender to determine precisely how the influence of curriculum differed across gender. Significant gender main effects indicate whether or not overall gender differences emerged (collapsed across curricula); we report these values, but they are not of primary concern in the current project.

2. Identifying and removing potential instructor effects

Because we had no baseline measure to use to statistically control for variation in instruction quality, our strategy was to identify and remove instructors who significantly impacted (positively or negatively) a given measure above and beyond the other variables included in

the model.¹ A dummy-coded dichotomous variable was created for each instructor (e.g., for the variable *Instructor A*, 1 = student had instructor A and 0 = student did not have instructor A). For each dependent variable and for each instructor, a hierarchical regression was conducted in which the dummy-coded instructor variable was added to the full regression model described above. If the instructor variable was a significant predictor (either positive or negative) in this model (i.e., the B value for the instructor term was significant at $p < 0.05$), we concluded that the given instructor had produced some instructor-specific influence beyond that of instruction type, and, accordingly, we excluded this instructor's students from the analysis for the given dependent measure. (Note that all descriptive statistics reported in this paper and appendixes also exclude these instructors' students.) Because this analysis reduced our sample size for certain analyses, and thus the power to detect effects for some measures, we consider this method to be a conservative approach to assessing the effects of Active Physics on student performance and CLASS attitudes.

III. RESULTS

A. Fall-only analyses

1. Does Active Physics improve conceptual learning relative to traditional-lecture physics?

Although c is our primary measure, it is worth noting that normalized change (c), post-FCI, and absolute gains all showed an advantage for Active Physics over traditional-lecture physics, as seen in Appendix B, Table IV. Figure 1 highlights the advantage of Active Physics for c , and our regression model confirmed that this advantage was significant (see Appendix B, Table V). The main effect of gender and the gender \times instruction-type interaction were not significant.

2. Does Active Physics improve attitudes relative to traditional-lecture physics?

As seen in Fig. 2 (see Appendix B, Table VI for detailed descriptive data), we examined the CLASS's overall score, the eight traditional factors [16], and two factors we identified through our own factor analysis (Learning Approach and Solving Approach). Across the Fall semester overall CLASS scores decreased by 2.5% among students in traditional-lecture physics and increased by 1.5% among students in Active Physics. Particularly striking is that *every* CLASS category showed a nominal negative shift across the semester in traditional-lecture physics. In contrast, Active

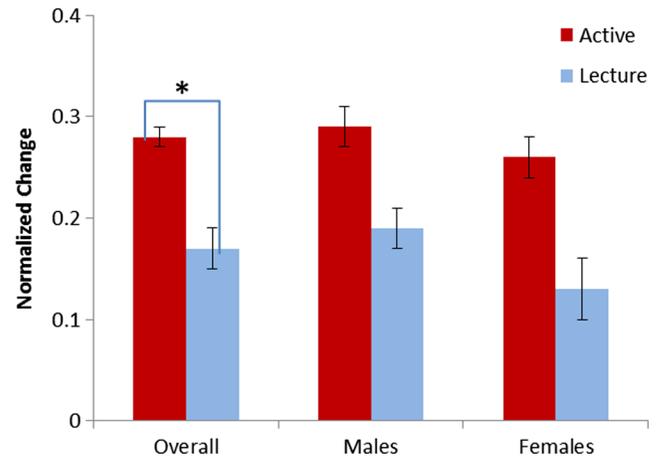


FIG. 1. Normalized change on the FCI across the Fall semester for students in Active Physics and traditional-lecture physics. Error bars represent the standard error of the mean. $*(p < 0.05)$ for instruction-type main effect in regression analyses. Significance tests were not conducted separately within each gender because the gender \times instruction-type effect was not significant.

Physics showed a nominal positive shift on 7 of the 10 categories, and for the 3 that decreased, the negative shift was smaller than in traditional-lecture physics.

Regression analyses were conducted separately for the overall CLASS score and for each CLASS category, and these analyses (see Table V) revealed significant instruction-type effects for the overall CLASS score, for 4 of the 8 CLASS categories typically used in the literature (Personal Interest, Real-World Connection, Problem-Solving General, and Problem-Solving Confidence), and for 1 of the 2 factors identified by our factor analysis (Solving Approach). In summary, the regression analyses confirmed that Active Physics conferred significant

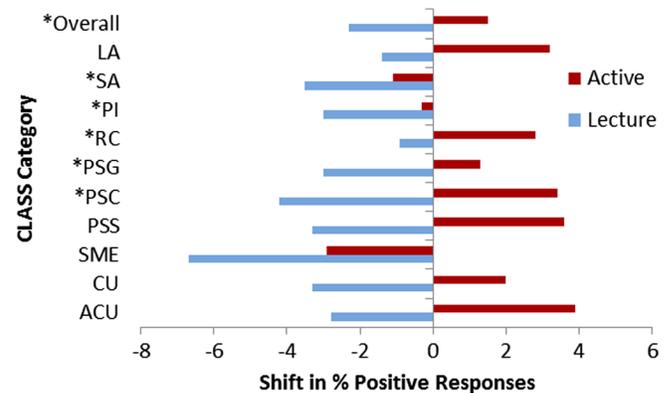


FIG. 2. CLASS Shifts from pre-Fall to post-Fall for students in Active Physics and traditional-lecture physics. $*(p < 0.05)$ for instruction-type main effects in regression analyses. LA, Learning Approach; SA, Solving Approach; PI, Personal Interest; RC, Real-World Connection; PSG, Problem Solving–General; PSC, Problem Solving–Confidence; PSS, Problem Solving Sophistication; SME, Sense Making/Effort; CU, Conceptual Understanding; ACU, Applied Conceptual Understanding.

¹We also attempted to represent all 10 instructors with 9 dichotomous dummy-coded variables and enter them all into the equation at once. We thought this would allow us to examine effects after removing all instructor-related variance. However, these models would not converge, leading us to the conservative but less elegant approach that we took.

advantages over traditional-lecture physics on both important attitudinal outcomes and conceptual understanding (i.e., the FCI).

3. Does the impact of Active Physics vary across gender?

The gender \times instruction-type term of the hierarchical regression analyses signaled whether the effect of instruction type was constant or variable across genders. The analyses revealed that the instruction-type \times gender term was nonsignificant for the FCI, for the overall CLASS score, and for 9 of the 10 CLASS subcategories (see Table V). Thus, the benefits of Active Physics emerged equally across the genders. Because of the general lack of significant gender \times instruction-type interactions, we had no statistical basis to motivate modeling the effect of instruction type separately within each gender and did not conduct these gender-specific analyses (with the exception of Personal Interest, which we address below). However, we do present the descriptive statistics for each gender separately in Fig. 1 (see Table IV for more detail) and Fig. 3 (see Appendix B, Table VII for more detail). Overall, these figures reveal comparable effects of Active Physics with males and females.

Personal Interest was the only CLASS variable for which the instruction-type \times gender interaction was significant. Examination of the means reveals that instruction type had little effect on males as shifts were negligible in both traditional-lecture physics and Active Physics. For females, however, Active Physics led to almost no shift in Personal Interest, but traditional-lecture physics led to nearly a 7% decrease. Hierarchical modeling analysis performed separately on each gender confirmed that the effect of instruction type was not significant among males ($B = -5.6$, $p = 0.25$), but it was significant among females ($B = 10.31$, $p < 0.001$).

4. Fall-only data summary

The Fall-only data demonstrate that (a) Active Physics (relative to traditional-lecture physics) provided significant benefits to both conceptual learning and attitudes and (b) the benefits were fairly consistent across genders.

B. Whole-year analysis

In order to address our final research question, whether the impact of Active Physics changes across the second semester, we conducted a second set of analyses on outcomes of the Spring semester: the BEMA (administered at the beginning and end of Spring) and the CLASS (with analyses using responses from pre-Fall and post-Spring surveys).

1. Does Active Physics improve conceptual learning relative to traditional-lecture physics?

As with the FCI, c was calculated for BEMA and used as the primary index of conceptual learning for the material

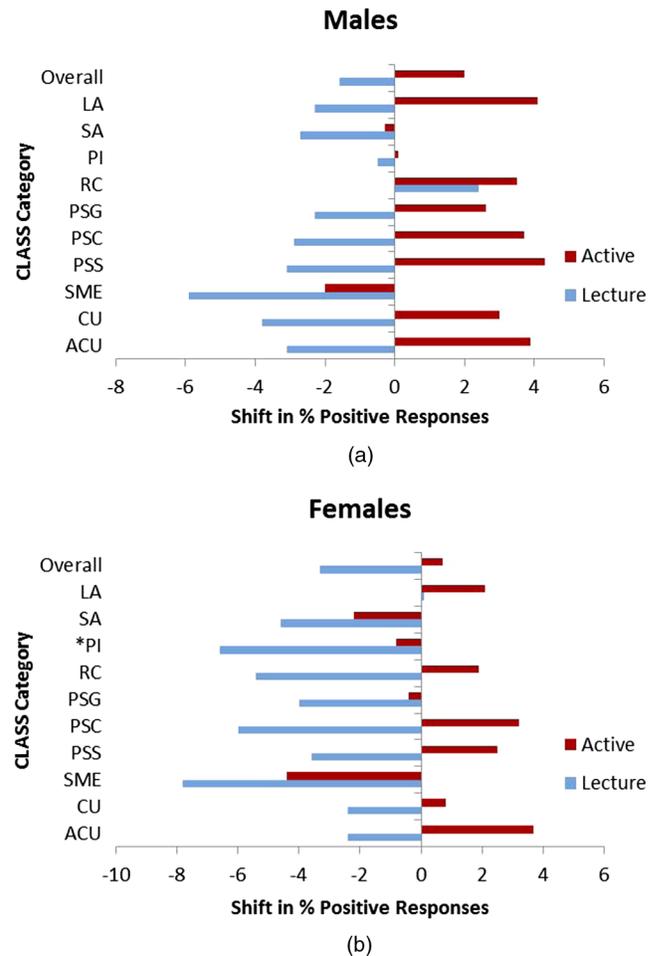


FIG. 3. CLASS shifts from pre-Fall to post-Fall for (a) males and (b) females in Active Physics and traditional-lecture physics. Significance tests were conducted separately within each gender only for PI, because of a significant gender \times instruction-type interaction effect. The instruction-type effect for PI was significant ($p < 0.05$) for females (indicated by *) but not for males.

covered in the spring semester. As seen in Fig. 4, the values of c were relatively small for both Active Physics and traditional-lecture physics (see Table IV for detailed descriptives). Regression analysis (see Appendix B, Table VIII) confirmed that the effect of instruction type on c was not significant (nor were the main effect of gender or the gender \times instruction-type interaction). Thus, our data reveal no advantages of Active Physics on conceptual learning in the Spring.

2. Does Active Physics improve attitudes relative to traditional-lecture physics?

Figure 5 (see Appendix B, Table IX for detailed descriptive data) shows that, unlike the Fall-only shifts, the whole-year shifts were negative for all CLASS indices for both Active Physics and traditional-lecture physics. However, for every CLASS index the shifts were less negative for Active Physics than for traditional-lecture physics. In fact,

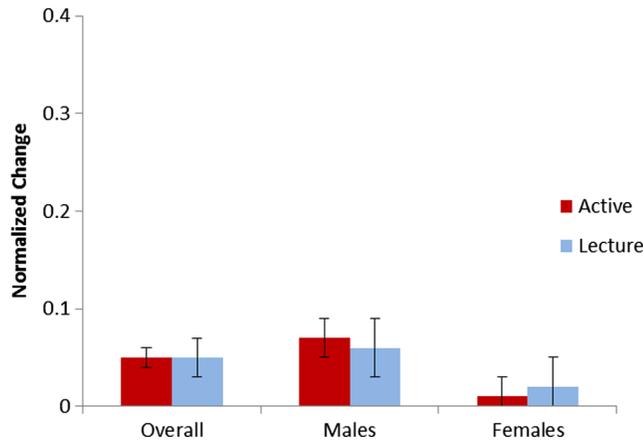


FIG. 4. Normalized change on the BEMA across the Spring semester for students in Active Physics and traditional-lecture physics. Error bars represent the standard error of the mean. Significance tests were not conducted separately within each gender because the gender \times instruction-type interaction effect was not significant.

regression analyses revealed that the benefits of Active Physics were actually more robust in the whole-year analysis (see Table VIII), as the overall CLASS score, one of our derived factors, and 7 of the 8 traditional CLASS subscales revealed a main effect of instruction type, indicating a significant advantage of Active Physics. Solving Approach was the only subscale to shift from significant in the Fall to nonsignificant across the whole year, while Learning Approach, Problem-Solving Sophistication, Conceptual Understanding, and Applied Conceptual Understanding all shifted from nonsignificance in the Fall to significance across the whole year.

3. Does the impact of Active Physics vary across gender?

In the hierarchical-modeling analysis, no variable demonstrated a significant gender \times instruction-type interaction, indicating that the benefits of Active Physics were consistent across genders. The interaction effect for Personal Interest, which was present at post-Fall, was no

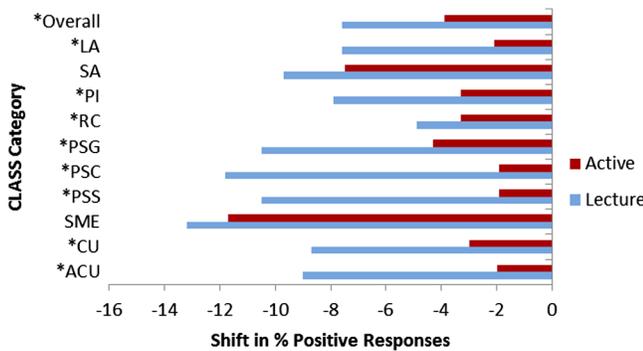


FIG. 5. CLASS shifts from pre-Fall to post-Spring for students in Active Physics and traditional-lecture physics. $*(p < 0.05)$ for instruction-type main effects in regression.

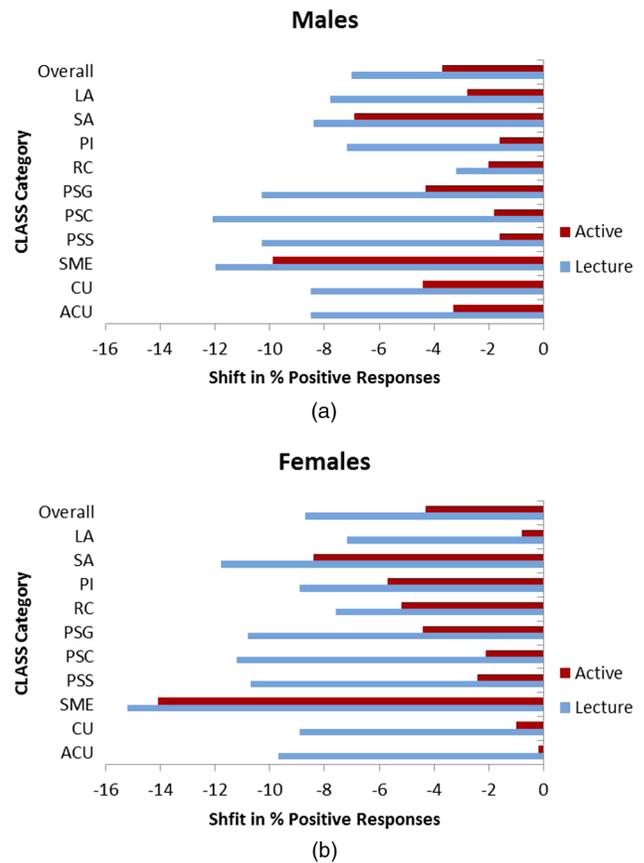


FIG. 6. CLASS shifts from pre-Fall to post-Spring for (a) males and (b) females. Significance tests were not conducted separately within each gender because no gender \times instruction-type interaction effects were significant in the whole-year analyses.

longer significant when analyzed across the whole year. In addition, the descriptive data in Fig. 6 (see Appendix B, Table X for detailed descriptive data) suggest that, if anything, the benefits of Active Physics (relative to traditional-lecture physics) were slightly greater for females than for males.

4. Relative to traditional-lecture physics, how does Active Physics impact attitudes across the whole year?

Although all CLASS indices trended toward an Active-Physics advantage in both the Fall-only and whole-year analysis, a comparison of the Fall-only (Fig. 2 and Table V) and whole-year (Fig. 5 and Table VIII) data indicates that despite the overall downward trajectory of CLASS attitudes in the Spring, additional exposure to Active Physics increased the reliability of these Active-Physics benefits (relative to traditional-lecture physics) for CLASS attitudes.

IV. DISCUSSION

A. Results summary and implications

Our primary aim in this project was to conduct a multiyear, multi-instructor evaluation of whether some of

the gains (particularly in attitudes) that have been reported recently with inquiry-based curricula could be achieved with a scalable IE curriculum. Our evaluation revealed that our Active-Physics curriculum, which is based on Moore's *Six Ideas That Shaped Physics*, produced benefits over traditional-lecture physics on both conceptual learning (the FCI in the Fall) and attitudes (CLASS shifts both during the Fall-only and across the whole year). Further, we attempted through our methodology and analytic strategy to have our results reflect the effects of the curriculum *per se* rather than other potential influences such as sample differences, instructor differences, or presemester differences on our outcome measures.

Especially encouraging is that these gains were achieved with a curriculum that has surmountable challenges to implementation. This curriculum was used in classes of up to 130 students at Washington University in St. Louis (the same size classes as traditional-lecture physics) and could likely be used in even larger classes without losing effectiveness. The class was held in a normal lecture hall and required no classroom redesign or technological enhancements. Many of the IE activities were taken or adapted from the *Six Ideas* or Mazur's Conceptests, easing the burden associated with developing materials for Active Physics. Also, Active Physics required the same staff resources as traditional-lecture physics. Despite the addition of daily homework in Active Physics, Active Physics TAs and traditional-lecture TAs had comparable workloads. This is because Active Physics TAs intentionally avoided giving detailed feedback, so that students would seek out answers on their own. Still, significant work was involved in transitioning to the new curriculum and incorporating the IE activities and the discussion cohesively into the lecture, potentially deterring instructors from adopting it. Encouragingly, the curriculum has been readily adopted by the faculty at Washington University. The majority of instructors at Washington University now use this curriculum, and faculty resistance was not a serious concern at this particular institution. In fact, currently only one section of traditional-lecture physics is offered, compared to five sections of Active Physics as of Fall 2013 (there will be six Active-Physics sections in Fall 2014). The single traditional-lecture section remains because a small minority of faculty and students still prefer it, and for scheduling reasons (traditional lecture can meet twice a week whereas Active Physics must meet three times a week). Instructors' overall willingness to adopt the Active-Physics curriculum suggests that this and similar curricula potentially could be implemented on a much wider scale than pure inquiry-based curricula, conferring benefits to many students who might otherwise experience no IE and receive none of the accompanying benefits.

B. Gender results

Another positive aspect of the Active-Physics curriculum is that its benefits are, for the most part, consistent across the

genders. The only instance in which the impact of Active Physics differed across gender was for the Personal Interest subscale of the CLASS in the Fall-only analyses. In this case, curriculum had a small impact on males, but Active Physics buffered females from the noticeable drop in Personal Interest reported by females in traditional-lecture physics. All other analyses supported the idea that Active Physics was equally beneficial for males and females.

1. A note about gender gaps

Although not our primary focus, a mention of gender gaps is warranted in light of previous research [10,22,23]. Within the Active-Physics curriculum, the gender gap decreased slightly on the FCI (from 21.9% at pretest to 18% at post-test) but increased slightly on the overall CLASS score in the Fall (from 7.5% to 9.2%). Similarly, across the whole year, the CLASS gender gap increased from 7.0 to 7.6. From these data alone, one might conclude that Active Physics increased the gender gap, at least on the CLASS. However, very similar changes in the gender gap occurred in the traditional-lecture section on the CLASS (a 0.1% female advantage at pre-Fall to a 1.1% male advantage at post-Fall in the Fall-only sample; a 1.1% female advantage at pre-Fall to a 0.6 male advantage at post-Spring), so Active Physics did not enhance the gender gap beyond what happened with a traditional-lecture baseline. More importantly, among females Active Physics (relative to traditional-lecture physics) produced consistently higher CLASS scores, so Active Physics clearly benefitted females despite the slight gender-gap increase occurring within this curriculum.

C. Spring results

Consistent with previous findings [4,24], CLASS scores decreased across the Spring semester in Active Physics (from 66.0 at post-Fall to 60.7 at post-Spring). These results in isolation give the impression that Active Physics negatively impacted attitudes. However, comparison with the traditional-lecture course reveals that despite the absolute decrease, Active Physics benefitted student CLASS attitudes relative to the traditional-lecture baseline. Perhaps students are unable, in one semester, to gain deep conceptual understanding of the more abstract and sometimes counterintuitive content such as electromagnetism, statistical mechanics, and quantum mechanics. Students may fill out the post-Spring CLASS with this material and their lack of conceptual understanding in mind, thus reporting less "expertlike" perceptions of the field, regardless of curriculum. However, Active Physics (and likely other scalable IE curricula) still had positive effects in the form of buffering students against the sharper decreases present in traditional-lecture physics. In fact, despite the overall CLASS decrease, the benefits of Active Physics relative to traditional-lecture physics were even more statistically robust after the entire two-semester sequence than after the Fall semester alone.

One unanticipated finding in the current study was that the Active-Physics group did not show significantly higher gains

on the BEMA than the traditional-lecture curriculum. In fact, the gains for both groups were negligible ($c \sim 0.05$ for both curricula). One explanation for this finding might be that at Washington University only a portion of the semester was spent covering electricity and magnetism (45% for Active Physics, 65% for traditional-lecture physics) and a significant amount of time passed between covering this material and taking the post-Spring BEMA (~ 7 – 8 weeks for Active Physics, ~ 4 – 5 weeks for traditional-lecture physics). This is unlike previous studies [4,24], in which an entire semester was devoted to electricity and magnetism, and may help explain the minimal gains for both curricula on the BEMA relative to those previous studies.

D. Student perceptions of Active Physics

Despite benefitting from Active Physics, students may nonetheless show resistance to aspects of the class that differ from traditional-lecture courses [31], and student resistance (whether real or perceived) may discourage instructors from implementing these techniques [18]. In order to explore actual student resistance, we examined student evaluations from both Active Physics and traditional-lecture physics. We directly compared these curricula on students' overall class ratings, perceived amount of learning, and perceived workload, and also examined individual comments from Active-Physics students to help determine student attitudes toward the curriculum. Relative to traditional-lecture physics, Active Physics produced significantly higher overall ratings (on a 1–7 scale) for both the Fall [5.77 vs 4.44; $t(1303) = 16.32$, $p < 0.001$] and Spring [5.72 vs 3.78; $t(972) = 19.21$, $p < 0.001$]. Active Physics also led students to perceive a significantly greater amount of learning in both the Fall [5.49 vs 4.72; $t(1301) = 8.59$, $p < 0.001$] and the Spring [5.60 vs 4.31; $t(970) = 12.06$, $p < 0.001$]. Student comments suggest that in-class activities (i.e., 2-minute problems) were a large part of both the overall positivity and the perception of greater learning in Active Physics. In fact, the most common theme in student comments was that in-class activities were both enjoyable and helpful for learning (e.g., “I enjoyed the interactivity of the classroom, especially how learning occurs through active involvement”).

Also promising is that, although Active Physics required students to complete daily homework on yet untaught topics, students in the two curricula had nearly identical perceptions of workload in both the Fall (4.94 for Active Physics vs 4.99 for traditional lecture) and Spring (5.15 for Active Physics vs 5.14 for traditional lecture). Despite completing objectively more homework, students in Active Physics did not perceive the workload to be any greater. This finding makes sense in light of a common theme in student comments in Active Physics—that the daily homework helped them to keep up with the material. (For example, “I ... liked the setup and the weekly and daily homework. It helped me keep up with the course

material without overloading me.”) The increased homework load may have eased the burden associated with studying for exams, thus creating a roughly equivalent workload (or at least perceived workload).

Some students did express the expected resistance toward Active Physics, particularly the task of completing homework prior to learning the material in class (e.g., “I ... must teach myself the material, making time in class useless”). However, these comments were balanced out by those expressing positive attitudes toward daily homework, and the student ratings illustrate a clear preference for the Active-Physics course relative to the traditional-lecture physics. These data demonstrate that instructors need not fear upsetting the class by introducing a curriculum based on IE. However, instructors should be prepared to deal with the reality that some students may resist the adoption of at least some aspects of a transformed curriculum.

E. What are the beneficial elements of Active Physics?

We have shown that the Active-Physics curriculum provides both conceptual learning and attitudinal benefits relative to the traditional-lecture curriculum. However, our data cannot disentangle which specific aspect(s) of Active Physics produced these positive effects. As shown in Table III, the Active-Physics curriculum differs from traditional-lecture physics in six discrete ways: textbook or materials, 2-minute problems, demonstration-centered small-group activities, problem-solving approach, daily homework, and homework revision. Isolating the effects of specific aspects of the curriculum is important practically because some instructors may be willing to introduce some but not all of these elements. For instance, some instructors may have no problem changing the homework but are uncomfortable adding in-class group work. Others may be willing to adopt in-class changes but do not want the extra workload of grading daily homework.

By pinpointing the aspects sufficient to produce benefits, a larger number of instructors may adopt these effective changes, and a larger number of students may reap the benefits. For example, if we were able to conclude that 2-minute problems alone produce these benefits, then instructors reluctant to change homework could still implement the 2-minute problems, benefitting their students. If, on the other hand, it turns out that all elements of the curriculum are necessary to produce these benefits, instructors would know that there is no reason to put in the effort to change some elements unless they are willing to adopt the new curriculum as a whole. With this in mind, future research should seek to isolate the aspects of the Active-Physics curriculum (and other curricula) that produce conceptual and attitudinal benefits, and also seek to develop new, (relatively) easily implementable interventions that potentially can benefit a large proportion of physics students at a wide range of institutions.

V. CONCLUSIONS

The current study provides strong evidence based on rigorous methods and analyses that a scalable IE curriculum, Active Physics, produced both conceptual learning and attitudinal benefits relative to what is often the only viable alternative—a traditional-lecture course. Even in the Spring semester, in which CLASS indices declined in an absolute sense, the presence of this traditional-lecture comparison allowed us to see that Active Physics conferred benefits in a relative sense. The current literature contains several examples of scalable IE courses leading to absolute declines in CLASS [10, 17], but our results highlight the possibility that even courses with somewhat “mixed” results may actually be producing consistent benefits relative to what students would experience in traditional-lecture courses. Another important finding is that the Active-Physics curriculum had clear benefits for both males and females across both semesters of introductory physics. Importantly, relative to pure inquiry-based curricula, this curriculum can be used effectively in large courses and has fewer practical obstacles to widespread implementation. Thus, Active Physics and other scalable IE curricula may allow the benefits of IE to reach a greater number of students than would otherwise be possible.

ACKNOWLEDGMENTS

We are grateful to the Washington University Department of Physics, especially all the instructors who allowed us to gather data from their courses. We are especially grateful to Tom Bernatowicz for his efforts in implementing the Active-Physics curriculum at Washington University. We also thank Frances Thuet for her help with data collection and management. This work was supported by HHMI under Grants No. 52005911 and No. 52006961.

APPENDIX A: CLASS ITEMS IN FACTORS 1 AND 2

Factor 1—Learning approach (rote versus conceptual)

- A significant problem in learning physics is being able to memorize all the information I need to know
- After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic
- Knowledge in physics consists of many disconnected topics
- When I solve a physics problem, I locate an equation that uses the variables given in the problem and plug in the values
- There is usually only one correct approach to solving a physics problem
- I cannot learn physics if the teacher does not explain things well in class
- I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations

- Understanding physics basically means being able to recall something you’ve read or been shown
- I do not spend more than five minutes stuck on a physics problem before giving up or seeking help from someone else
- If I don’t remember a particular equation needed to solve a problem on an exam, there’s nothing much I can do (legally!) to come up with it.
- If I want to apply a method used for solving one physics problem to another problem, the problems must involve very similar situations
- I can usually figure out a way to solve physics problems.
- If I get stuck on a physics problem, there is no chance I’ll figure it out on my own.

Factor 2—Solving approach (algorithmic versus concept based)

- When I am solving a physics problem, I try to decide what would be a reasonable value for the answer
- I am not satisfied until I understand why something works the way it does
- I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations
- If I get stuck on a physics problem my first try, I usually try to figure out a different way that works
- In doing a physics problem, if my calculation gives a result very different from what I’d expect, I’d trust the calculation rather than going back through the problem
- In physics, it is important for me to make sense out of formulas before I can use them correctly
- In physics, mathematical formulas express meaningful relationships among measurable quantities
- To learn physics, I only need to memorize solution to sample problems
- Spending a lot of time understanding where formulas come from is a waste of time
- There are times I solve a physics problem more than one way to help my understanding
- To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed
- When I solve a physics problem, I explicitly think about which physics ideas apply to the problem
- When studying physics, I relate the important information to what I already know rather than just memorizing it the way it is presented

APPENDIX B: SUPPLEMENTARY DATA

Tables IV–X present additional details about the data presented in Figs. 1–6. Tables IV, VI, VII, IX, and X provide detailed descriptive statistics, and Tables V and VIII provide statistics from the hierarchical regression analyses.

TABLE IV. Detailed descriptive statistics for the FCI and BEMA. SE indicates Standard error.

Force Concept Inventory								
	Lecture				Active			
	<i>N</i>	Pre-FCI (SE)	Post-FCI (SE)	<i>c</i> (SE)	<i>N</i>	Pre-FCI (SE)	Post-FCI (SE)	<i>c</i> (SE)
Overall	366	63.3 (1.1)	68.0 (1.1)	0.17 (0.02)	773	65.0 (0.8)	73.6 (0.8)	0.28 (0.01)
Male	222	66.7 (1.4)	71.3 (1.4)	0.19 (0.02)	444	74.2 (0.9)	80.9 (0.9)	0.29 (0.02)
Female	144	58.1 (1.6)	63.0 (1.6)	0.13 (0.03)	329	52.6 (1.1)	63.8 (1.2)	0.26 (0.02)

Brief Electricity and Magnetism Assessment								
	Lecture				Active			
	<i>N</i>	Pre-BEMA (SE)	Post-BEMA (SE)	<i>c</i> (SE)	<i>N</i>	Pre-BEMA (SE)	Post-BEMA (SE)	<i>c</i> (SE)
Overall	279	30.7 (0.8)	37.5 (1.1)	0.05 (0.02)	738	32.9 (0.6)	38.9 (0.2)	0.05 (0.01)
Male	165	32.3 (1.2)	39.6 (1.6)	0.06 (0.03)	437	36.0 (0.8)	42.7 (1.0)	0.07 (0.02)
Female	114	28.2 (1.1)	34.4 (1.6)	0.02 (0.03)	301	28.5 (0.8)	33.5 (1.0)	0.01 (0.02)

TABLE V. Unstandardized regression coefficients for the Fall-only CLASS and FCI. Bold cells indicate a significant effect ($p < 0.05$).

	Instruction type		Gender		Instruction type \times gender	
	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>
FCI	16.23	0.02	5.53	0.27	-5.37	0.39
Overall	7.35	0.002	5.05	0.19	-2.76	0.51
LA	2.32	0.42	-1.34	0.63	3.40	0.38
SA	8.85	0.03	6.34	0.19	-3.11	0.54
PI	11.45	<0.001	12.93	<0.001	-10.88	0.02
RC	13.78	0.003	14.80	0.002	-9.66	0.07
PSG	9.53	0.04	9.48	0.02	-3.49	0.45
PSC	16.15	0.001	15.37	0.02	-8.08	0.24
PSS	14.48	0.051	11.63	0.22	-6.66	0.52
SME	3.72	0.19	2.00	0.48	2.36	0.47
CU	1.95	0.54	-1.78	0.54	8.38	0.05
ACU	6.69	0.13	3.04	0.56	-0.49	0.93

TABLE VI. Detailed descriptive statistics for the Fall-only CLASS. SE indicates Standard error.

	Lecture				Active			
	<i>N</i>	Pre (SE)	Post (SE)	Shift	<i>N</i>	Pre (SE)	Post (SE)	Shift
Overall	282	61.0 (1.0)	58.7 (1.1)	-2.3	763	64.5 (0.6)	66.0 (0.7)	1.5
LA	303	54.2 (1.3)	52.8 (1.3)	-1.4	298	59.2 (1.2)	62.4 (1.3)	3.2
SA	282	71.1 (1.1)	67.6 (1.3)	-3.5	763	74.0 (0.7)	72.9 (0.8)	-1.1
PI	357	55.3 (1.5)	52.3 (1.6)	-3.0	374	61.1 (1.5)	60.8 (1.6)	-0.3
RC	282	64.4 (1.8)	63.5 (1.9)	-0.9	763	70.5 (1.1)	73.3 (1.1)	2.8
PSG	282	65.9 (1.5)	62.9 (1.6)	-3.0	374	69.1 (1.3)	70.4 (1.3)	1.3
PSC	282	64.8 (1.8)	60.6 (2.0)	-4.2	700	70.3 (1.1)	73.7 (1.1)	3.4
PSS	282	53.1 (1.9)	49.8 (1.9)	-3.3	687	61.8 (1.2)	65.4 (1.1)	3.6
SME	357	69.8 (1.2)	63.1 (1.4)	-6.7	763	71.6 (0.8)	68.7 (0.9)	-2.9
CU	357	60.7 (1.4)	57.4 (1.5)	-3.3	304	63.8 (1.6)	65.8 (1.7)	2.0
ACU	357	50.2 (1.4)	47.4 (1.4)	-2.8	763	57.6 (0.9)	61.5 (0.9)	3.9

TABLE VII. Detailed descriptive statistics among males and females for the Fall-only CLASS. SE indicates Standard error.

		Lecture				Active			
		<i>N</i>	Pre (SE)	Post (SE)	Shift	<i>N</i>	Pre (SE)	Post (SE)	Shift
Overall	Male	163	61.1 (1.3)	59.5 (1.4)	-1.6	441	67.6 (0.7)	69.6 (0.8)	2.0
	Female	119	60.8 (1.4)	57.5 (1.6)	-3.3	322	60.3 (1.0)	61.0 (1.1)	0.7
LA	Male	180	54.4 (1.7)	52.1 (1.8)	-2.3	166	62.7 (1.6)	66.8 (1.7)	4.1
	Female	123	53.8 (1.9)	53.9 (1.9)	0.1	132	54.8 (1.8)	56.9 (2.0)	2.1
SA	Male	163	70.2 (1.5)	67.5 (1.8)	-2.7	441	76.1 (0.8)	75.8 (0.9)	-0.3
	Female	119	72.5 (1.6)	67.9 (2.0)	-4.6	322	71.2 (1.2)	69.0 (1.3)	-2.2
PI	Male	213	57.1 (2.1)	56.6 (2.1)	-0.5	212	64.8 (2.0)	64.9 (2.0)	0.1
	Female	144	52.5 (2.3)	45.9 (2.4)	-6.6	162	56.3 (2.4)	55.5 (2.5)	-0.8
RC	Male	163	65.3 (2.3)	67.7 (2.3)	2.4	441	73.3 (1.4)	76.8 (1.3)	3.5
	Female	119	63.1 (3.0)	57.7 (3.0)	-5.4	322	66.7 (1.7)	68.6 (1.8)	1.9
PSG	Male	163	66.3 (2.0)	64.0 (2.0)	-2.3	212	72.9 (1.6)	75.5 (1.6)	2.6
	Female	119	65.4 (2.2)	61.4 (2.5)	-4.0	162	64.2 (2.1)	63.8 (2.1)	-0.4
PSC	Male	163	64.1 (2.4)	61.2 (2.5)	-2.9	404	76.1 (1.3)	79.8 (1.3)	3.7
	Female	119	65.8 (2.8)	59.8 (3.2)	-6.0	296	62.3 (1.9)	65.5 (1.8)	3.2
PSS	Male	163	54.4 (2.4)	51.3 (2.5)	-3.1	395	68.2 (1.4)	72.5 (1.4)	4.3
	Female	119	51.3 (2.9)	47.7 (2.7)	-3.6	292	53.2 (1.8)	55.7 (1.8)	2.5
SME	Male	213	68.9 (1.6)	63.0 (1.8)	-5.9	441	73.8 (1.0)	71.8 (1.2)	-2.0
	Female	144	71.0 (1.7)	63.2 (2.1)	-7.8	322	68.7 (1.3)	64.3 (1.5)	-4.4
CU	Male	213	60.0 (1.9)	56.2 (2.0)	-3.8	166	67.4 (2.1)	70.4 (2.2)	3.0
	Female	144	61.6 (2.2)	59.2 (2.2)	-2.4	138	59.5 (2.4)	60.3 (2.4)	0.8
ACU	Male	213	50.1 (1.9)	47.0 (1.9)	-3.1	441	62.1 (1.2)	66.0 (1.2)	3.9
	Female	144	50.4 (2.2)	48.0 (2.1)	-2.4	322	51.5 (1.5)	55.2 (1.5)	3.7

TABLE VIII. Unstandardized regression coefficients for the whole-year CLASS and BEMA. Bold cells indicate a significant effect ($p < 0.05$).

	Instruction type		Gender		Instruction type \times gender	
	<i>B</i>	<i>p</i>	<i>B</i>	<i>p</i>	<i>B</i>	<i>P</i>
BEMA	-1.47	0.70	2.41	0.58	2.04	0.68
Overall	10.40	<0.001	6.38	0.07	-5.50	0.15
LA	16.01	<0.001	7.39	0.16	-8.34	0.13
SA	6.41	0.16	5.47	0.26	-3.22	0.53
PI	16.24	0.003	12.46	0.04	-6.04	0.35
RC	11.25	0.02	9.04	0.08	-4.75	0.41
PSG	14.57	<0.001	7.16	0.14	-3.83	0.47
PSC	15.38	0.002	3.68	0.50	2.99	0.63
PSS	29.02	<0.001	17.85	0.07	-13.79	0.17
SME	2.44	0.55	4.39	0.34	1.66	0.74
CU	19.87	0.02	10.69	0.13	-12.53	0.09
ACU	17.06	0.004	5.95	0.28	-6.75	0.25

TABLE IX. Detailed descriptive statistics for the whole-year CLASS. SE indicates Standard error.

	Lecture				Active			
	<i>N</i>	Pre (SE)	Post (SE)	Shift	<i>N</i>	Pre (SE)	Post (SE)	Shift
Overall	250	60.6 (1.1)	53.0 (1.2)	-7.6	671	64.6 (0.6)	60.7 (0.8)	-3.9
LA	250	55.3 (1.4)	47.7 (1.5)	-7.6	671	61.9 (0.8)	59.8 (1.0)	-2.1
SA	250	70.9 (1.2)	61.2 (1.6)	-9.7	616	74.1 (0.8)	66.6 (1.0)	-7.5
PI	250	56.3 (1.8)	48.4 (2.0)	-7.9	671	63.1 (1.1)	59.8 (1.3)	-3.3
RC	250	62.7 (2.0)	57.8 (2.1)	-4.9	671	70.8 (1.1)	67.5 (1.3)	-3.3
PSG	250	66.5 (1.5)	56.0 (1.7)	-10.5	671	71.6 (1.0)	67.3 (1.1)	-4.3
PSC	250	66.2 (1.8)	54.4 (2.0)	-11.8	588	70.7 (1.3)	68.8 (1.3)	-1.9
PSS	250	52.5 (2.0)	42.0 (1.9)	-10.5	588	62.7 (1.3)	60.8 (1.3)	-1.9
SME	250	70.4 (1.5)	57.2 (1.8)	-13.2	616	72.1 (0.9)	60.4 (1.1)	-11.7
CU	250	60.4 (1.8)	51.7 (1.9)	-8.7	671	68.7 (1.0)	65.7 (1.1)	-3.0
ACU	250	50.6 (1.7)	41.6 (1.7)	-9.0	671	58.0 (1.0)	56.0 (1.1)	-2.0

TABLE X. Detailed descriptive statistics among males and females for the whole-year CLASS. SE indicates Standard error.

		Lecture				Active			
		<i>N</i>	Pre (SE)	Post (SE)	Shift	<i>N</i>	Pre (SE)	Post (SE)	Shift
Overall	Male	155	60.2 (1.4)	53.2 (1.6)	-7.0	396	67.5 (0.8)	63.8 (1.0)	-3.7
	Female	95	61.3 (1.6)	52.6 (1.9)	-8.7	275	60.5 (1.0)	56.2 (1.3)	-4.3
LA	Male	155	55.1 (1.8)	47.3 (1.9)	-7.8	396	65.4 (1.1)	62.6 (1.3)	-2.8
	Female	95	55.7 (2.2)	48.5 (2.3)	-7.2	275	56.7 (1.3)	55.9 (1.4)	-0.8
SA	Male	155	69.6 (1.6)	61.2 (2.0)	-8.4	359	76.3 (0.9)	69.4 (1.2)	-6.9
	Female	95	73.0 (1.9)	61.2 (2.5)	-11.8	257	71.1 (1.3)	62.7 (1.6)	-8.4
PI	Male	155	58.2 (2.4)	51.0 (2.8)	-7.2	396	67.8 (1.4)	66.2 (1.5)	-1.6
	Female	95	53.0 (2.8)	44.1 (2.9)	-8.9	275	56.3 (1.8)	50.6 (2.0)	-5.7
RC	Male	155	61.9 (2.5)	58.7 (2.6)	-3.2	396	73.9 (1.5)	71.9 (1.5)	-2.0
	Female	95	64.0 (3.2)	56.4 (3.4)	-7.6	275	66.4 (1.8)	61.2 (2.2)	-5.2
PSG	Male	155	66.9 (2.0)	56.6 (2.1)	-10.3	396	75.8 (1.1)	71.5 (1.3)	-4.3
	Female	95	65.9 (2.3)	55.1 (2.8)	-10.8	275	65.6 (1.6)	61.2 (1.8)	-4.4
PSC	Male	155	65.6 (2.4)	53.5 (2.5)	-12.1	355	75.8 (1.5)	74.0 (1.6)	-1.8
	Female	95	67.1 (2.9)	55.9 (3.5)	-11.2	233	62.9 (2.1)	60.8 (2.2)	-2.1
PSS	Male	155	53.4 (2.6)	43.1 (2.4)	-10.3	355	68.1 (1.5)	66.5 (1.6)	-1.6
	Female	95	50.9 (3.1)	40.2 (2.9)	-10.7	233	54.5 (2.1)	52.1 (1.9)	-2.4
SME	Male	155	69.4 (1.8)	57.4 (2.3)	-12.0	359	74.0 (1.1)	64.1 (1.3)	-9.9
	Female	95	72.0 (2.4)	56.8 (2.9)	-15.2	257	69.4 (1.5)	55.3 (1.8)	-14.1
CU	Male	155	59.7 (2.3)	51.2 (2.4)	-8.5	396	72.2 (1.3)	67.8 (1.5)	-4.4
	Female	95	61.5 (2.8)	52.6 (3.0)	-8.9	275	63.6 (1.6)	62.6 (1.7)	-1.0
ACU	Male	155	49.8 (2.2)	41.3 (2.1)	-8.5	396	62.0 (1.3)	58.7 (1.4)	-3.3
	Female	95	51.8 (2.7)	42.1 (2.7)	-9.7	275	52.2 (1.6)	52.0 (1.5)	-0.2

- [1] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [2] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, *Science* **332**, 862 (2011).
- [3] K. K. Perkins, W. K. Adams, S. J. Pollock, N. D. Finkelstein, and C. E. Wieman, Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey, *AIP Conf. Proc.* **790**, 61 (2005).
- [4] S. J. Pollock, Transferring transformations: Learning gains, student attitudes, and the impacts of multiple instructors in large lecture courses, *AIP Conf. Proc.* **818**, 141 (2006).
- [5] E. Brewster, L. Kramer, and G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS, *Phys. Rev. ST Phys. Educ. Res.* **5**, 013102 (2009).
- [6] E. Brewster, A. Traxler, J. de la Garza, and L. Kramer, Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020116 (2013).
- [7] V. K. Otero and K. E. Gray, Attitudinal gains across multiple universities using the Physics and Everyday Thinking curriculum, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020104 (2008).
- [8] B. A. Lindsey, L. Hsu, H. Sadaghiani, J. W. Taylor, and K. Cummings, Positive attitudinal shifts with the Physics by Inquiry curriculum across multiple implementations, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010102 (2012).
- [9] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, in *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), see review at <http://www.per-central.org/document/ServeFile.cfm?ID=4517>.
- [10] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **5**, 010101 (2009).
- [11] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, *Am. J. Phys.* **74**, 118 (2006).
- [12] T. Moore, *Six Ideas That Shaped Physics* (McGraw-Hill, New York, 2007), 2nd ed.
- [13] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985).
- [14] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [15] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- [16] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
- [17] J. de la Garza and H. Alarcon, Assessing students' attitudes in a college physics course in Mexico, *AIP Conf. Proc.* **1289**, 129 (2010).
- [18] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020102 (2007).
- [19] M. Dancy and C. Henderson, Pedagogical practices and instructional change of physics faculty, *Am. J. Phys.* **78**, 1056 (2010).
- [20] M. Milner-Bolotin, T. Antimirova, A. Noack, and A. Petrov, Attitudes about science and conceptual physics learning in university introductory physics courses, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020107 (2011).
- [21] E. F. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum, *Am. J. Phys.* **77**, 629 (2009).
- [22] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010107 (2007).
- [23] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020121 (2013).
- [24] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a "smog of bias," *Phys. Rev. ST Phys. Educ. Res.* **6**, 020112 (2010).
- [25] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- [26] N. J. Dorans, "Correspondences between ACT and SAT I Scores," College Board Report No. 99-1, 1999.
- [27] H. D. Young and R. A. Freedman, *Sears and Zemansky's University Physics with Modern Physics* (Pearson, New York, 2007), 12th ed.
- [28] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer instruction: Engaging students one-on-one, all at once, *Research-Based Reform of University Physics* **1**, 40 (2007).
- [29] J. D. Marx and K. Cummings, Normalized change, *Am. J. Phys.* **75**, 87 (2007).
- [30] M. J. Cahill, J. Zhao, M. Repice, R. F. Frey, and M. A. McDaniel, Evaluation of alternative factor structures for the Colorado Learning Attitudes about Science Survey (to be published).
- [31] R. M. Felder and R. Brent, Navigating the bumpy road to student-centered instruction, *Coll. Teach.* **44**, 43 (1996).