Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs

Ping Zhang,[1,*] Lin Ding,[2,†] and Eric Mazur[3]

[1]*Department of Physics, Beijing Normal University, Beijing 100875, People's Republic of China*
[2]*Department of Teaching and Learning, The Ohio State University, Columbus, Ohio 43210, USA*
[3]*Department of Physics and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 4 September 2015; published 17 January 2017)

This paper analyzes pre-post matched gains in the epistemological views of science students taking the introductory physics course at Beijing Normal University (BNU) in China. In this study we examined the attitudes and beliefs of science majors ($n = 441$) in four classes, one taught using traditional (lecture) teaching methods, and the other three taught with Peer Instruction (PI). In two of the PI classes, student peer groups were constantly changing throughout the semester, while in the other PI class student groups remained fixed for the duration of the semester. The results of the pre- and post-test using the Colorado Learning Attitudes about Science Survey showed that students in traditional lecture settings became significantly more novicelike in their beliefs about physics and learning physics over the course of a semester, a result consistent with what was reported in the literature. However, all three of the classes taught using the PI method improved student attitudes and beliefs about physics and learning physics. In the PI class with fixed peer groups, students exhibited a greater positive shift in attitudes and beliefs than in the other PI class with changing peer groups. The study also looked at gender differences in student learning attitudes. Gender results revealed that female science majors in the PI classes achieved a greater positive shift in attitudes and beliefs after instruction than did male students.

DOI: 10.1103/PhysRevPhysEducRes.113.010104

I. INTRODUCTION

Much research has been conducted to uncover students' attitudes and beliefs in undergraduate physics courses, most of which reports students' attitudes and beliefs about physics and learning physics deteriorate throughout the course of a semester of introductory instruction [1,2]. This means that as they progress through introductory courses, students' views do not become more like those of experts but instead become more novicelike. These patterns have been consistently uncovered in nearly all traditional lectures and even in many reformed courses [1–4].

There is mounting evidence, however, that positive shifts in student attitudes and beliefs about physics and learning physics have been shown to be achievable by employing some reformed pedagogies. In one of the first studies showing positive attitudinal gains related to alternative curricula, Otero and Gray [5] found that students taught with the *Physics and Everyday Thinking (PET)* curriculum [6,7] experienced expertlike shifts on the Colorado Learning Attitudes about Science Survey (CLASS) [1] after instruction. Research in modeling instruction [8], which was designed to involve students in scientific practices such as model building and revision, also resulted in positive shifts in student epistemologies over the course of a semester [9,10]. Recently, Lindsey and colleagues [9,10] found that preservice teachers experienced positive shifts on the CLASS after instruction in courses utilizing the *Physics by Inquiry* (*PBI*) curriculum [11].

Similar outcomes have also been reported in other instructional settings. For instance, by implementing a new introductory physics course designed specifically for life sciences, Crouch and colleagues [12] found an increase in student attitudes about the subject matter measured by CLASS after a semester of instruction. Interestingly, even in a traditional class, Milner-Bolotin and colleagues [13] observed a positive change in student learning attitudes. They further reported that such a positive change appeared to be closely related to student prior exposure to high school physics.

In light of these findings, as well as those from traditionally taught classes, Madsen and colleagues [14] conducted a meta-analysis of 24 studies that used CLASS to gauge learners' attitudinal views about physics before and after formal instruction. Based on the collective data, the

*zhangping@bnu.edu.cn
†Ding.65@osu.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.

researchers proposed four factors that could possibly relate to learners' attitudinal gains. These were teaching method, class size, student population, and preinstructional scores. Although these factors did not emerge as definitive outcomes from the meta-analysis, the researchers cogently argued that the lack of homogeneity (or in other words, the unbalanced nature) of the available data from the published literature could possibly obscure the findings. This indeed is a major challenge in conducting meta-analysis, where the results are directly bound by the quality of the studies under analysis. To this end, controlled empirical studies with higher affordances of homogeneity can better explicate factors that are associated with learners' gains in epistemological views.

In this study, we chose to focus on Chinese college students taking pedagogically reformed calculus-based physics classes using Peer Instruction (PI) at Beijing Normal University—a comprehensive large research institution in China. A uniqueness of this student population is the relative uniformity [15] of their racial backgrounds, preuniversity preparation, and class types to which they were exposed (see Sec. III for details). More importantly, we followed the quasiexperimental design, while also taking into consideration the students' baseline data (pretest scores), to compare PI students with their counterparts in a traditional class of similar size and the same physics course (see Sec. III for details). As such, we controlled all of the aforementioned factors of teaching method, class size, student population, and pretest scores to systematically investigate the effects of PI instruction on students' learning attitudes.

Also novel of the current study is our attempt to examine the effects of peer grouping (fixed or variable groups) on students' learning attitudes. Specifically, we implemented PI teaching in two classes of the same physics course; one with fixed student peer groups and the other with variable peer groups (see Sec. III for details). By controlling the above factors, we compared the two classes to examine the effects of grouping on students' attitudinal changes—a topic that has not been studied but can add to our understandings about the mechanisms of interactive teaching and its effect on student learning attitudes.

Specifically, in this study we address the following questions:

(1) To what extent do students' attitudes and beliefs change after a semester of instruction in a traditionally taught class and in the PI class?
(2) How do students in these two types of classes compare in terms of their change in attitudes toward physics and learning physics over the course of a semester?
(3) In PI classes, how do learning attitudes of the students participating in fixed peer groups compare to those of the students in variable peer groups?
(4) Are there any gender differences in attitudes and beliefs between PI and traditional classes?

II. RESEARCH BACKGROUND

Epistemologies, also known as attitudes, beliefs, and views about knowledge and learning, have been studied in physics education research for the past several decades [16–21]. In particular, this research has focused on shifting students' views about the subject toward those aligned to experts. Novicelike beliefs are characterized by viewing physics knowledge as disconnected facts, and this often results in rote memorization of content knowledge. These differ drastically from the beliefs of experts and scientists, who view the subject as a matrix of coherent, connected topics and use reasoning through one's knowledge base to figure out problems or questions. As previously mentioned, research into students' attitudes and beliefs has revealed a trend of negative shifts over the course of a semester of introductory instruction [1,2]. However, several more recent studies with reformed courses using different pedagogical techniques indicate that there can be expertlike shifts in student epistemologies [5,9,11]. The research reported here investigates shifts in learning attitudes among students taking classes using the Peer Instruction (PI) method [22].

Peer Instruction was developed by the physics education research group at Harvard University and was designed to engage students in active, peer-led discussions in order to help learners solidify conceptual understanding and encourage them to learn from one another much as they will when they become members a scientific community [22–25]. In this course, class time is used for learners to construct concepts rather than for instructors to lecture. Differing from those in traditional lectures, PI students are required to read and complete assignments before attending class. Thus, the instructional time can center on a set of conceptual questions called ConcepTest [22], which students are held accountable for answering by using what they have gained through the preinstructional reading assignments. In class, the instructor poses ConceptTest questions and the students choose their own answers, first without consulting their peers [18]. Students report their answers either through a classroom response system (clickers) [26,27] or other low-tech devices (e.g., flashcards or raising their hands). If more than a third of the class choose an incorrect answer, students then discuss their answers with others sitting around them. After discussion, students will revote and the instructor will discuss the question with the class as a whole, explain the answer, and then move on to the next point [22]. Each key point in a lecture takes an average of minimum of 15 min, including 7–10 min of minilecturing and 5–8 min for ConcpTest questions. Therefore, a 1 hr class typically requires about four key points. Interested readers can see Ref. [18] for PI and ConceptTest sample questions.

Although scholarly publications on the design and implementation of PI pedagogies have been abundant, there is a dearth of work dedicated specifically to evaluating

the impact of the PI method on student epistemological views. In fact, to the best of our knowledge, little has been published on this particular topic. Also, a whole host of literature has invariably pointed out that unless explicitly catering to the learners' epistemological development, even reformed pedagogies will likely fail to bring up students' correct views about the nature of physics and the nature of learning physics. Peer Instruction, as one of many reformed pedagogical methods, although focusing on student active learning, does not, in principle, contain explicit instruction on topics about the nature of physics or the nature of learning physics. Therefore, it is of great interest and need that we study how the PI method will bring about students' epistemological changes. Moreover, in this study we specifically used the PI method with Chinese students—an under-researched population in the literature of peer instruction—to study how students in the PI classes compare to their peers in epistemological views about physics and learning physics.

III. METHODS

A. Context and participants

Participants were first-year undergraduate science majors at BNU. Students admitted to the science majors at BNU had taken high-school physics and met the high proficiency requirements for sciences and mathematics determined by their scores on the National University Entrance Examinations. All the participants in the study were from four large classes (each with the size of approximately 110 students or above) of the same physics course, namely, the second-semester calculus-based introductory physics. Of these four classes, one (class A) was traditionally taught and the other three (classes B, C, and D) adopted the PI methods. All classes were delivered by the same instructor who had significant experience with PI and had been teaching calculus-based introductory physics for decades. Note that unlike in the large U.S. higher institutions where students of the same major can choose to attend different classes (or sessions) of the same physics course, at BNU science students from the same major are assigned to the same physics class. In our sample, students in two of the classes (classes A and C) were geology majors, and students in the other two (classes B and D) were chemistry majors. Classes A and B were taught in the autumn semester of 2011, and classes C and D were taught in the autumn semester of 2012. This fact notwithstanding, the textbook and course syllabus were the same for all the classes.

At BNU, we implemented the PI methods in several ways. First, as suggested by Mazur [22], students were asked to read the textbook and answer a few questions online prior to class. Upon completion of these questions, students would receive a small amount of credit for their responses. To further incentivize student reading before class, quizzes were administered using a classroom response system (clickers) at the beginning of each class. All of our implementations of PI included student peer discussion during class. However, in two of the classes (classes B and C) the discussion groups changed for every session, and in the other class (class D) discussion groups were selected by students at the beginning and kept fixed for the rest of the semester. As a summary, Table I shows the information of each class and participants.

In order to examine the effects of PI on students' attitudes and beliefs about physics, we selected class A as a control group and class B as an experimental group. In Class A, the teaching method was traditional (i.e., pure lecture). In class B, the teaching method was PI with variable student groups. This means that students were constantly forming new peer discussion groups with those who happened to sit near them in each class session. To verify the results, we studied two additional classes in the following year. Specifically we compared class C as a second experimental group against the control group class A, because both consisted of students from the same major of geology. This allowed us to have a better control of student background. In addition, we compared class D with class B, as both were chemistry majors using PI instruction. However, a crucial difference between them was that in class D students self-selected peer groups of 3–4 members that remained fixed throughout the entire semester, whereas the peer groups in class B were changing for each session.

B. Instrument

To measure students' attitudes and beliefs about the nature of physics and learning physics, we used the Colorado Learning Attitudes about Science Survey [1]. The CLASS instrument is useful for students at all levels of physics and is appropriate for both physics majors and nonmajors. Student responses to the 42-Likert item statements in the survey are scored along a novice-expert continuum and then averaged as favorable (aligning to

TABLE I. Participants and demographics for each class included in the study [N (Female, Male)].

Class	Instructional method	Major	Total enrollment	Matched responses
A	Traditional	Geology	114 (79, 35)	94 (69, 25)
B	Peer Instruction (with variable groups)	Chemistry	147 (116, 31)	134 (106, 28)
C	Peer Instruction (with variable groups)	Geology	106 (70, 36)	95 (62, 33)
D	Peer Instruction(with fixed groups)	Chemistry	138 (111, 27)	118 (94, 24)

experts' views) or unfavorable (aligning to novicelike beliefs). Responses are then scored overall and in eight categories. These eight categories are (i) students' ability to apply physics concepts (applied conceptual understanding), (ii) their understanding of physics concepts (conceptual understanding), (iii) their level of interest in the subject (personal interest), (iv) how relevant physics is to their everyday lives (real world connection), (v) how difficult students perceive it is to understand physics concepts (sense making or effort), and three problem solving categories, which are (vi) the participant's confidence in their ability to solve problems (PS confidence), (vii) general problem solving issues (PS general), and (viii) how students solve complex physics problems (PS sophistication).

In this study we used the published Mandarin CLASS, which was translated from the original English version and validated through both quantitative and qualitative techniques. Here, we provide a brief summary of the validation work, and details can be found in Ref. [28]. First, we translated the English CLASS into Chinese. This Chinese version was then back translated into English by a bilingual education researcher who had no knowledge about the instrument at that time. Next, we compared the back-translated CLASS with the original survey to further fine-tune our Chinese translation until all divergences were resolved. Additionally, we conducted student interviews to ensure that our translation maximally preserved the original meaning intended by the CLASS designers, and that students did not misconstrue the survey questions. Moreover, we calculated the consistency level of the instrument in our study (reliability alpha $= 0.89$), and found that it was above the acceptable level ≥ 0.70 for group measurement [29–31].

We administered the Mandarin CLASS to 441 science major students in four classes of the first-year introductory physics course. In all cases, the instrument was administered as an in-class, paper-and-pencil survey, and the students were given a maximum of 20 min to complete the task. Although no incentives or penalties were offered, students were reminded to provide their honest responses. In order for a student's results to be included for analysis, the student must have completed both the pre- and post-instruction surveys.

IV. DATA ANALYSIS AND RESULTS

Pre-post data were matched and analyzed at the individual student level by using the template provided by the developers of CLASS [1]. These spreadsheets utilize a scoring scheme that calculates average percent favorable scores (responses that align to experts') for the overall performance and performance on the eight categories established by the authors. For the purpose of our research goals, we used within-sample t tests for the measurement of students' pre-post gains for each class. Additionally, we used between-sample t tests for comparisons between the different classes, including those between PI and traditional classes, between fixed and variable student PI groups, and between male and female students. Note that because t tests *by nature* are robust to violations of the normality assumption and also because our samples are of large size, the use of within- and between-sample t tests, as was done in the current study, is valid both theoretically and practically [29,32]. That said, we paid particular attention to the homogeneity of our data relating to the equal variance assumption, a condition that could possibly affect the accuracy of the exact t statistic values [32]. Therefore, in cases where the homogeneity assumption failed to hold, we used the unequal-variance t statistic and its corresponding p error rate and effect size to make inferences.

Specifically, we focused on comparisons between classes of (1) traditional instruction versus PI instruction (2) PI with variable peer groups versus PI with fixed peer Additionally, we investigated whether or not there were gender differences for traditional and PI classes.

A. Comparison of PI and traditional class

We first used within-sample paired t tests to determine if significant changes occurred in all of the classes between the pre- and post-instruction surveys. Overall, students in class A experienced a significant negative shift in their attitudes and beliefs over the course of a semester (see Table II). On the other hand, classes B, C, and D all experienced significant positive changes overall (see Table II). In all cases, the effect size (ES) based on the standard Cohen's d [33] was nontrivial (>0.2), indicating that the changes we detected, be they positive or negative, held a practical significance (in addition to a statistical

TABLE II. Overall pre-and postinstruction percent favorable responses (± standard error) for the four classes. Shifts reported in bold text indicate significance ($p < 0.05$); paired sample t tests are included.

					t test		
Class		Pre	Post	Shift	t	p	ES
A	(Traditional)	56.9 (1.6)	52.5 (1.6)	**−4.4(1.4)**	**−3.17**	**0.002**	**0.30**
B	(PI with variable groups)	50.0 (1.2)	53.2 (1.2)	**3.2(1.1)**	**2.86**	**0.005**	**0.22**
C	(PI with variable groups)	54.2 (1.4)	57.0 (1.4)	**2.8(1.3)**	**2.15**	**0.034**	**0.21**
D	(PI with fixed groups)	51.5 (1.2)	57.0 (1.1)	**5.5(1.2)**	**4.44**	**<0.001**	**0.44**

significance). Table II summarizes the overall differences for all four classes.

A closer look at the individual categories provides more details on the contrast between the PI and the traditional class. In class A, which utilized traditional lectures as the delivery method, the students shifted toward more novice-like beliefs in all of the CLASS categories (see Table III). Among them, three categories exhibited significant negative shifts: problem solving sophistication, conceptual understanding, and applied conceptual understanding (see Table III). Conversely, in classes B, C and D that followed the PI pedagogy, the shifts in all but one category in the class C were positive (see Table III). Moreover, the PI students achieved a statistically significant positive shift on the categories of personal interest (in classes B, C, and D), problem solving general (in classes C and D), and real world connection (in class B) (see Table III). All these positive shifts showed a nontrivial effect size. In other words, the peer discussion approach appeared to have aided students' interest, views about problem solving, and understandings of physics in relation to the real world. The only category on which students experienced a decrease in favorable responses after instruction was real world connection (in class C), though this difference was not significant [$t(94) = -1.68$, $p = 0.097$].

In addition, we paid particular attention to classes A and C, because students in both classes were of the same major—geology.

Overall, students in class A had a higher percentage of favorable responses before instruction (Table II) than did class C. However, this preinstruction difference was insignificant [$t(187) = -1.29$, $p = 0.198$], thereby suggesting that both classes were comparable in their attitudes and beliefs about physics prior to the course. The overall postinstruction scores reversed this trend, with class A having a lower percentage of favorable responses than class C (Table II). This difference was significant and of medium size [$t(184) = 2.11$, $p = 0.036$, ES $= 0.31$], indicating that the PI class demonstrated more expertlike views about physics than the traditional class.

B. Differences between class B (PI-variable groups) and class D (PI-fixed groups)

Both classes consisted of students from the same major (chemistry) and were taught using the PI approach. However, the student groups were randomly formed in class B but kept the same for the entire semester in class D.

Class B exhibited a slightly lower overall percentage of favorable scores before instruction than did class D (Table II). However, this difference was insignificant

TABLE III. Scores in % favorable response (± standard error) of each individual category of the pre and post-CLASS test for classes A, B, C, and D. Shifts in bold text indicate significance ($p < 0.05$). Paired sample t tests are included.

	Class A (tradition)						Class B (PI-variable groups)					
				t test						t-test		
Categories	Pre	Post	Shift	t	p	ES	Pre	Post	Shift	t	p	ES
Personal interest	48.9(2.5)	46.2(2.4)	−2.7(2.3)	−1.16	0.250	···	39.4(2.0)	50.4(2.2)	**11.0(2.1)**	**5.20**	**<0.001**	**0.40**
Real world connection	71.0(2.9)	70.1(3.1)	−0.9(3.2)	−0.28	0.784	···	61.6(2.8)	67.9(2.5)	**6.3(2.7)**	**2.35**	**0.020**	**0.21**
PS general	56.3(2.3)	53.3(2.2)	−3.0(2.3)	−1.29	0.199	···	53.5(1.9)	57.2(1.7)	3.8(2.0)	1.92	0.058	···
PS confidence	47.6(3.2)	46.5(2.9)	−1.1(3.1)	−0.34	0.735	···	47.2(2.5)	50.6(2.5)	3.4(2.8)	1.18	0.241	···
PS sophistication	39.0(2.9)	34.3(2.7)	**−4.8(2.2)**	**−2.18**	**0.032**	0.18	34.3(2.2)	36.1(2.4)	1.8(2.3)	0.79	0.430	···
Sense making or effort	65.7(2.1)	61.6(2.0)	−4.1(2.4)	−1.65	0.102	···	58.9(1.9)	62.8(1.8)	3.9(2.1)	1.18	0.073	···
Conceptual understanding	64.8(2.1)	57.8(2.3)	**−7.0(2.2)**	**−3.20**	**0.002**	**0.33**	55.9(2.1)	58.2(1.8)	2.3(2.2)	1.04	0.298	···
Applied conceptual understanding	45.6(2.3)	39.4(2.4)	**−6.2(2.1)**	**−3.00**	**0.003**	**0.27**	36.9(2.1)	38.4(2.0)	1.4(2.0)	0.72	0.473	···

	Class C (PI-variable groups)						Class D (PI-fixed groups)					
				t test						t-test		
Categories	Pre	Post	Shift	t	p	ES	Pre	Post	Shift	t	p	ES
Personal interest	46.7(2.4)	53.7(2.4)	**7.0(2.7)**	**2.60**	**0.011**	**0.30**	43.0(2.1)	55.2(2.0)	**12.2(2.6)**	**4.76**	**<0.001**	**0.55**
Real world connection	73.2(2.6)	68.2(2.9)	−5.0(3.0)	−1.68	0.097	···	68.6(2.8)	71.3(2.5)	2.7(3.1)	0.86	0.39	···
PS general	53.0(2.1)	58.4(2.2)	**5.4(2.4)**	**2.21**	**0.030**	**0.25**	52.3(2.0)	58.1(1.8)	**5.8(2.3)**	**2.51**	**0.013**	**0.28**
PS confidence	47.4(2.7)	49.5(3.2)	2.1(3.3)	0.63	0.532	···	44.9(2.7)	50.1(2.5)	5.2(3.1)	1.68	0.095	···
PS sophistication	34.2(2.3)	38.4(2.6)	4.2(2.5)	1.68	0.096	···	37.2(2.4)	37.4(2.1)	0.2(2.5)	0.09	0.930	···
Sense making or effort	60.2(2.4)	65.0(2.3)	4.8(2.6)	1.83	0.070	···	59.8(2.0)	63.0(1.8)	3.1(2.3)	1.38	0.169	···
Conceptual understanding	60.7(2.0)	63.3(2.1)	2.6(2.2)	1.17	0.245	···	58.6(2.1)	59.9(1.7)	1.3(2.4)	0.52	0.605	···
Applied conceptual understanding	41.9(2.2)	44.2(2.3)	2.3(2.3)	1.00	0.323	···	39.8(1.9)	40.5(2.0)	0.7(2.4)	0.30	0.766	···

[$t\ (250) = 0.830,\ p = 0.407$], meaning that the pretest performance on CLASS for the students in the two classes was comparable. However, the post-instruction scores indicated that class B exhibited a lower overall percentage of favorable scores than class D (Table II), a difference that was statistically significant and of medium size [$t(248) = 2.317,\ p = 0.021,\ \mathrm{ES} = 0.29$]. In other words, students in the PI class with fixed discussion groups showed more expertlike views about physics than their counterparts in the PI class with variable groups.

C. Gender differences over the course of the semester

According to the overall CLASS scores for all of our participants, females appeared less expertlike in their attitudinal views than males at the beginning of the course (female: average $= 51.5\% \pm 0.7\%$ standard error; male: average $= 56.7\% \pm 1.5\%$). The overall difference was statistically significant with a nontrivial effect size [$t\ (439) = 3.38,\ p = 0.001,\ \mathrm{ES} = 0.37$]. Table IV shows the gender breakdown in the overall results for the traditional class A and each of PI classes (classes B, C, and D). As seen, both females and males in class A experienced a significant and a medium-sized negative shift in their attitudes and beliefs during a semester (see Table IV). In the PI classes the male students overall showed some slight changes, but none was significant (see Table IV). Interestingly, the female students in the PI classes displayed a significant overall positive shift after instruction, and such a shift was nontrivial (see Table IV).

To further evaluate the impact of PI teaching on gender, we calculated both genders' performances on CLASS for all the PI classes combined. As shown in Table V, the shifts for the male students on most categories, regardless of the directions, were insignificant ($p > 0.40$). Only on the category of Personal Interest was there a significant difference, and this difference was a medium-sized, positive shift (Table V). As for the females, the overall result after a semester of PI instruction revealed a positive shift that was statistically significant and of a medium size (Table V). Specifically, three categories were statistically significant with a nontrivial effect size, including: personal interest, problem solving general, and senses making or effort (Table V).

To understand the above differences, we randomly selected from each of the three PI classes (B, C, and D) six student groups consisting of both genders, and audio recorded their peer discussions on five ConcepTest

TABLE IV. Differences in overall percent favorable responses and overall shifts between males and females in all four classes ($\pm$ standard error). Shifts in bold font indicate significance ($p < 0.05$). Paired sample t tests (t statistics, p values, and effect sizes—ES) are included.

	Females						Males					
				t test						t test		
Class	Pre	Post	Shift	t	p	ES	Pre	Post	Shift	t	p	ES
A	55.8(1.8)	52.4(1.9)	**−3.4(1.6)**	**−2.14**	**0.036**	**0.23**	60.0(3.3)	53.0(3.1)	**−7.0(2.7)**	**−2.58**	**0.017**	**0.43**
B	48.8(1.3)	53.1(1.4)	**4.3(1.2)**	**3.52**	**0.001**	**0.31**	54.7(3.5)	53.6(2.4)	−1.1(2.4)	0.47	0.643	···
C	52.8(1.5)	55.9(1.6)	**3.1(1.5)**	**2.00**	**0.040**	**0.25**	56.9(2.7)	59.1(2.8)	2.2(2.3)	0.93	0.360	···
D	50.4(1.4)	56.6(1.2)	**6.2(1.4)**	**4.41**	**<0.001**	**0.48**	55.6(2.6)	58.5(2.1)	2.9 (2.6)	1.09	0.287	···

TABLE V. Pre- and postinstruction differences in individual categories ($\pm$ standard error) in the PI group. Bold font represents significant differences ($p < 0.05$). Paired sample t tests (t statistics, p values, and effect sizes—ES) are included.

	Females						Males					
				t test						t test		
Categories	Pre	Post	Shift	t	p	ES	Pre	Post	Shift	t	p	ES
Overall	50.3(0.8)	55.0(0.8)	**4.7(0.8)**	**5.89**	**<0.000**	**0.35**	55.8(1.7)	57.1(2.1)	1.3(1.4)	0.91	0.368	···
Personal Interest	39.2(1.4)	51.1(1.5)	**11.9(1.6)**	**7.30**	**<0.001**	**0.51**	52.9(2.6)	58.4(2.3)	**5.5(2.7)**	**2.01**	**0.048**	**0.25**
Real World Connection	64.8(1.9)	67.9(1.8)	3.1(2.0)	1.52	0.129	···	74.4(3.0)	72.9(2.9)	−1.5(2.9)	−0.50	0.618	···
PS General	51.1(1.3)	57.2(1.3)	**6.0(1.5)**	**4.10**	**0.001**	**0.29**	58.5(2.4)	60.0(2.0)	1.5(2.6)	0.57	0.571	···
PS Confidence	44.0(1.7)	49.0(1.8)	**5.0(2.1)**	**2.37**	**0.019**	0.17	54.1(2.9)	53.7(2.9)	−0.4(3.3)	−0.12	0.905	···
PS Sophistication	33.4(1.5)	36.6(1.5)	**3.2(1.6)**	**2.00**	**0.047**	0.13	40.8(3.0)	38.8(2.8)	−2.0(2.9)	0.70	0.489	···
Senses Making/Effort	58.3(1.3)	63.1(1.3)	**4.9(1.5)**	**3.31**	**0.001**	**0.23**	63.5(2.5)	64.4(2.4)	0.8(3.0)	0.28	0.783	···
Conceptual Understanding	57.9(1.4)	60.7(1.2)	2.8(1.5)	1.81	0.071	···	58.9(2.6)	58.6(2.2)	−0.3(2.7)	−0.10	0.919	···
Applied Conceptual Understanding	38.3(1.3)	40.7(1.4)	2.4(1.5)	1.60	0.110	···	42.4(2.5)	40.8(2.5)	−1.6(2.4)	−0.65	0.518	···

questions. By listening to their verbal utterances and examining the total time lengths that each student took to articulate their thoughts, we calculated the average minutes used by each gender during their group discussion. We found that each female student in average engaged in verbal discussions for 59.1 ± 1.3 sec on every question, whereas each male student averagely participated in verbal discussion on every question for only 46.9 ± 1.5 sec. The difference between the two genders was statistically significant [$t(178) = 6.22$, $p < 0.001$], indicating that female students tended to speak up more during group discussion than their male counterparts.

V. DISCUSSION

A. Comparisons of traditional and PI instruction

The results that students in the traditional lecture (Class A) became more novicelike in their attitudes and beliefs about physics over the course of a semester are not surprising and align with the results from other research on this topic. However, what is unusual of our results in light of the existing large body of literature is that all three of the PI classes have consistently shown significant expertlike changes in student views about physics and learning physics. This finding adds credence to the small—but growing—body of research that reports increased expertlike responses in epistemologies when alternative instructional techniques are utilized in introductory physics courses.

In our study, there was no instructor effect that could confound the results, as all the four classes were taught by the same instructor who was experienced and equally dedicated to teaching both PI and traditional classes. However, our data were collected from two academic years (classes A and B in fall 2011, and classes C and D in fall 2012), so this might potentially confound our findings. That said, we put forth the following facts to argue that the student populations between the two years are comparable, and therefore the likelihood of such a conflation in our results is low.

First, the overall national ranking of BNU remained the same over the years and, hence, has likely attracted and recruited students with similar academic backgrounds during the two years of our work. Also, regardless of the year, all students who are admitted to BNU must take the national University Entrance Examination and exceed a certain level of performance. Because the content and difficulty level of the exams are controlled and kept stable by the Chinese Ministry of Education, and also because the admission requirements at BNU (particularly of the two majors under investigation in our study) have not changed, it is hardly the case that the incoming student populations in the two consecutive years would differ drastically. Moreover, unlike the U.S. education system, Chinese K-12 education is closely regulated and monitored by the federal government; therefore, there is a strong sense of uniformity and stability in school curricula, instruction, and assessment across different regions and over time. Since the students of interest in our study were across a short time period of two consecutive years, there is no reason to believe that drastic changes in school education would occur to lead to significant differences in students' attitudes toward learning physics. In fact, our findings regarding the little variation in student preinstruction performance on the CLASS (see Sec. IV B–C) are strong evidence to support this argument.

In looking at the results between the three PI classes, we need to keep in mind that classes B and C were taught by using the same instructional methods, while class D deviated from the typical PI format. In both classes B and C, the students in each session formed peer discussion groups with their neighboring peers, so the groups were variable from day to day. Conversely, we designed a different situation in class D where students chose group members at the beginning of the semester and were then required to stay in the same group throughout the entire semester. It is possible that staying in the same discussion group for the duration of the semester would allow students to build a strong rapport with each other, facilitating more open and honest discussions and, hence, conceivably increasing their overall gains in the epistemological views by the end of the semester. In fact, many of those in the same groups in class D were also roommates sharing the same dormitories. So, there was likely a stronger bond between the group members that might have led to more out-of-class interactions. This could positively contribute to the students' views about physics and learning physics. On the other hand, the group members in class B were constantly changing. Although students in that class might become acquaintances at the end of the semester, it was unlikely that they would reach the same level of togetherness or have the same strong sense of community as those in class D. Future work may need to more systematically tease out the influence of interlearner closeness on their learning.

Moreover, we chose to study class A against class C separately, because both were geology majors. This allowed us to reduce the confounding factor of student background in comparing traditional lecture-based instruction with PI instruction. The pretest results showed that there was no significant difference in the overall performance between the two classes, meaning that students in these two classes were roughly equivalent in their attitudinal views about physics before taking the course. The posttest results, however, showed that class C had a higher overall percentage of favorable response than did class A, hence suggesting that PI in fact improved students' attitudes and beliefs about the nature of physics and learning physics more than the traditional instruction.

B. Comparison of class B (PI-changing groups) and class D (PI-fixed groups)

We chose to compare class B with class D separately, because they were chemistry majors, which again allowed us a better control over students' background. In our study, both classes B and D were taught using PI methods. However, students in class B formed different discussion groups in each class session, and students in class D were kept in fixed groups for the entire semester. While class B showed a slightly lower overall percentage of favorable responses before instruction, the difference between them was insignificant. This means that students' prior views about the nature of physics and learning physics were comparable. However, after instruction students in class B exhibited a lower overall percentage of favorable responses than those in class D, and such a difference was found to be significant (see Sec. IV B). Combining these results, we came to conclude that students in class B did not improve as much in their epistemologies as did students in class D. In other words, when controlling for student background (same science major taking the same calculus-based physics course), class size (large classes of over 100 students taught by the same instructor), and pretest scores (comparable preinstructional CLASS performance), students who were engaged in the fixed peer discussion groups appeared to improve their attitudinal views more than did those engaged in variable peer groups.

C. Gender differences

In the traditionally taught class A, both female and male students demonstrated a significant decrease in their epistemological views after a semester of introductory physics instruction (cf. Table IV). Conversely, the three PI classes by and large experienced positive shifts on the overall CLASS performance after instruction. Looking at the data more closely, we see that males in all of the three PI classes (B, C, and D) did not exhibit significant overall shifts after instruction (Table IV). However, female students in all the PI classes exhibited statistically significant increase on the overall CLASS scores after instruction. This suggested that the PI instruction positively affected the female students' attitudes and beliefs about physics and learning physics, whereas male students' did not appear to be affected as much. These results add to the existing literature on gender difference by speaking directly to student attitudinal changes in the PI context [34,35]. Besides, in our study we recorded students' in-class discussions. Recordings revealed that females spoke more than their male counterparts, which could support the differences we saw between the genders, as females were more active than their male peers and thus could get more out of peer discussions. Note that our sample consisted of more females ($n = 331$) than males ($n = 110$). This could potentially skew our results. Further work through focused group interviews perhaps is useful to extract rich information regarding the underlying mechanisms of gender differences.

VI. CONCLUSIONS

In reference to traditionally taught physics, our results support previous findings, indicating that students' attitudes and beliefs about physics shifted toward more novicelike after a semester of introductory instruction [1,2,17,36]. The fact that students did not improve in their attitudes and beliefs following a semester of instruction is therefore not surprising. However, our results in the classes taught using PI instruction are a different story.

There is research to support the notion that different instructional techniques may improve students' epistemologies about the nature of physics and learning physics over a semester of introductory instruction. This has been found mostly in modeling instruction for preservice teachers [9] and in courses using the *Physics by Inquiry* curriculum [11] and *Physics and Everyday Thinking* curriculum [5]. Adding to the existing literature, our findings showed that PI, another alternate instructional technique, can afford more benefits in student attitudes and beliefs than does the traditionally taught course. In fact, this has been consistently reoccurring in our study, where we performed a series of controlled comparisons between different classes. Specifically, we considered and controlled several key factors that, according to the literature, could potentially influence the changes in students' views about physics. These included teaching method, student background, class size, and pretest scores. Invariably, students in the PI classes exhibited positive shifts in their learning attitudes measured by CLASS, whereas those in the traditional class showed negative shifts.

Also encouraging from our study is that female students seemed to benefit from peer instruction more than male students in terms of their attitudinal changes. One note worth taking is that through controlled comparisons our study directly probed the effect of student grouping on learners' attitudinal change in PI classes—a topic that has not been investigated. To this end, our study pushes toward a better understanding of how to optimize inactive teaching to further improve learners' epistemological sophistication. In the long run, the results of our study can be suggestive of possible instructional techniques and pedagogies that may change how we teach introductory physics students.

ACKNOWLEDGMENTS

This work is supported by 2015 Comprehensive Discipline Construction Fund of Faculty of Education, Beijing Normal University (No. 2015GJXM002). We also thank Julie Schell, Kelly Miller, and Laura Tucker at Harvard University and Ying Liang at Beijing Normal University for their intellectual discussion and assistance. Additionally, we are grateful that Dr. Katherine Mollohan at The Ohio State University has provided valuable comments to the manuscript.

[1] 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).

[2] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).

[3] M. Sahin, Effects of problem-based learning on university students' epistemological beliefs about physics and physics learning and conceptual understanding of Newtonian mechanics, J. Sci. Educ. Technol. **19**, 266 (2010).

[4] L. Ding and P. Zhang, Making of epistemologically sophisticated physics teachers: A cross-sequential study of epistemological progression from preservice to in-service teachers, Phys. Rev. Phys. Educ. Res. **12**, 020137 (2016).

[5] V. K. Otero and K. Gray, Attitudinal gains across multiple universities using the Physics and Everyday Thinking curriculum, Phys. Rev. ST Phys. Educ. Res. **4**, 020104 (2008).

[6] F. Goldberg, S. Robinson, and V. K. Otero, *Physics for Elementary Teachers* (It's About Time, Armonk, NY, 2006).

[7] F. Goldberg, S. Robinson, and V. K. Otero, *Physical Science and Everyday Thinking* (It's About Time, Armonk, NY, 2008).

[8] J. Jackson, L. Dukerich, and D. Hestenes, Modeling Instruction: An effective model for science education, Sci. Educat. **17**, 10 (2008).

[9] E. Brewe, L. H. 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).

[10] E. Brewe, A. Traxler, J. d. l. 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).

[11] B. Lindsey, L. Hsu, H. Sadaghiani, J. 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).

[12] C. H. Crouch, W. Panchompoo, and K. A. Renninger, Initial interest, goals, and changes in CLASS scores in introductory physics for life sciences, in *2013 PERC Proceedings, Portland*, edited by P. Engelhardt, A. Churukian, and D. Jones (American Association of Physics Teachers, College Park, 2013), pp. 105–108.

[13] 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).

[14] A. Madsen, S. B. 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).

[15] L. Ding, X. Wei, and K. Mollohan, Does Higher Education Improve Student Scientific Reasoning Skills?, Int. J. Sci. Math. Educ. **14**, 619 (2016).

[16] K. Gray, W. Adams, C. Wieman, and K. Perkins, Students know what physicists believe, but they don't agree: A study using the CLASS survey, Phys. Rev. ST Phys. Educ. Res. **4**, 020106 (2008).

[17] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. **73**, 372 (2005).

[18] B. Hofer and P. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, Rev. Educ. Res. **67**, 88 (1997).

[19] D. Hammer, Two approaches to learning physics, Phys. Teach. **27**, 664 (1989).

[20] D. Hammer, Students' beliefs about conceptual knowledge in introductory physics, Int. J. Sci. Educ. **16**, 385 (1994).

[21] D. Hammer, Epistemological Beliefs in Introductory Physics, Cognit. Instr. **12**, 151 (1994).

[22] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).

[23] C. H. Crouch, A. P. Fagen, J. P. Callan, and E. Mazur, Classroom demonstrations: Learning tools or entertainment?, Am. J. Phys. **72**, 835 (2004).

[24] A. P. Fagen, C. H. Crouch, and E. Mazur, Peer Instruction: Results from a Range of Classrooms, Phys. Teach. **40**, 206 (2002).

[25] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[26] L. Ding and N. W. Reay, Teaching undergraduate introductory physics with an innovative research-based clicker methodology, in *Research Based Undergraduate Science Teaching*, edited by D. Sunal, C. Sunal, E. Wright, C. Mason, and D. Zollman (Information Age Publishing, Charlotte, NC, 2014), Vol. 6.

[27] A. Lee, L. Ding, N. W. Reay, and L. Bao, Single-concept clicker question sequences, Phys. Teach. **49**, 385 (2011).

[28] P. Zhang and L. Ding, Large-scale survey of Chinese precollege students' epistemological beliefs about physics: A progression or a regression?, Phys. Rev. ST Phys. Educ. Res. **9**, 010110 (2013).

[29] L. Ding and X. Liu, Getting started with quantitative methods in physics education research, in *Getting Started in PER–Reviews in PER*, edited by C. Henderson and K. Harper (American Association of Physics Teachers College Park, MD, 2012), Vol. 2, pp. 1–33.

[30] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).

[31] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).

[32] A. Agresti and B. Finlay, *Statistical Methods for the Social Sciences*, 4th ed. (Pearson Education, Upper Saddle River, NJ, 2009).

[33] J. Cohen, A power primer, Psychol. Bull. **112**, 155 (1992).

[34] L. Kost, S. Pollock, and N. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[35] A. Madsen, A. Madsen, 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).

[36] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, J. Learn. Sci. **12**, 53 (2003).