

Evaluating peer discussion in calculus-based mechanics using a cluster-randomized crossover design

Takashi Umeda^{*}

Graduate School of Humanities and Social Sciences, *Hiroshima University*,
1-1-1 Kagamiyama, Higashihiroshima, Hiroshima 739-8524, Japan

Shuji Munejiri

Graduate School of Advanced Science and Engineering, *Hiroshima University*,
1-7-1 Kagamiyama, Higashihiroshima, Hiroshima 739-8521, Japan



(Received 23 March 2026; accepted 26 June 2026; published 14 July 2026)

Peer discussion (PD) is widely used as an active-learning strategy in physics education, yet most prior research has focused on conceptual multiple-choice questions. Much less is known about how peer discussion functions when applied to mathematically structured problems typical of calculus-based mechanics courses. To address this gap, we compared the effects of PD and instructor explanation (IE) during short in-class exercise sessions in a calculus-based introductory mechanics course ($N = 87$). Using a cluster-randomized crossover design with two course sections, instruction alternated between PD and IE across weekly exercise topics. After each lesson, students completed a 16-item post-test; 12 of these items were also administered on the final exam. Across the 16 post-test items, we found no statistically significant evidence that PD outperformed IE on any single item. In unadjusted item-level analyses, four items that required greater abstraction and multistep mathematical reasoning showed a trend favoring IE, although none of these differences remained statistically significant after Holm correction for multiple comparisons. On the final exam, no statistically significant differences were observed between conditions. These results suggest that, within the instructional conditions of this study, PD did not produce consistent performance gains relative to IE, and any item-level differences may depend on task characteristics such as mathematical abstraction, novelty, and multistep reasoning demands.

DOI: [10.1103/wjhq-n354](https://doi.org/10.1103/wjhq-n354)

I. INTRODUCTION

A. STEM education reform and challenges in university physics

In recent years, the need for reform in Science, Technology, Engineering, and Mathematics (STEM) education at universities has been widely recognized worldwide. Traditional lecture-based instruction has long served as an efficient way to deliver advanced content, but many studies have shown that one-way lectures have limited effectiveness for promoting deep understanding in general STEM education [1–3]. Reflecting these worldwide initiatives, national policies in Japan—such as those guided by the Central Council for Education—have strongly encouraged the shift toward active learning [4].

Physics is a discipline in which mathematical formalism and conceptual understanding are closely connected. Students must connect mathematical expressions with the physical ideas they represent, and conceptual understanding depends on this integration of mathematics and physics. As a result, passive listening alone often does not lead to strong learning. Approaches that promote active thinking have therefore been emphasized for many years [5,6]. Within physics education research (PER), this motivation has led to the development of many research-based instructional strategies (RBIS), whose effectiveness has been widely demonstrated [7–9].

Despite this strong evidence, the adoption of RBIS in university physics courses has not progressed smoothly. Previous studies report that instructors recognize the limitations of lecture-based teaching, know about RBIS, and often express willingness to adopt new methods. Yet many do not implement RBIS due to factors such as limited preparation time, classroom constraints, student resistance, insufficient instructor knowledge, and restrictions posed by syllabi [10–12]. Even when RBIS are adopted, they are frequently modified, and many instructors later discontinue their use [13,14]. These difficulties

^{*}Contact author: tumeda@hiroshima-u.ac.jp

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

highlight the importance of examining RBIS that can realistically be implemented under the actual constraints of university physics courses.

B. Constraints in calculus-based mechanics and the role of peer instruction

This study focuses on calculus-based introductory mechanics, which typically covers fundamental yet mathematically demanding topics such as vector kinematics, Newton's laws, oscillations described by differential equations, work and energy, conservation laws, rotational dynamics, gravitation, and motion in inertial and non-inertial frames. These topics involve substantial mathematical formalism—including differential equations, vector calculus, and line integrals related to work and energy—and must all be addressed within a limited number of class hours. As a result, instructors generally have little flexibility to remove content or add new activities. This structural constraint creates challenges for adopting certain forms of RBIS.

Among RBIS, peer instruction (PI) is one of the most influential and widely studied methods [15,16]. PI places central emphasis on the use of *ConceptTests*—multiple-choice conceptual questions designed by Mazur to target well-documented student difficulties and to elicit diverse responses. Designing effective *ConceptTests* typically relies on prior research and classroom experience to identify common student difficulties and to construct questions that use them productively [17]. *ConceptTests* are thus crucial for stimulating productive peer discussion (PD). However, in calculus-based mechanics, students must not only confront deeply ingrained conceptual difficulties—which are well documented in foundational PER—but also master complex and entirely new mathematical formalisms. Integrating conceptual understanding with procedural mathematical fluency significantly increases students' cognitive load. Because the curriculum is exceptionally dense, instructors often feel pressured to prioritize traditional problem-solving practice to ensure mathematical competency. Consequently, they cannot easily introduce additional, purely conceptual *ConceptTests* without sacrificing essential course content.

For these reasons, full implementation of PI, including dedicated *ConceptTests*, is often not feasible in this course context. When instructors seek to integrate elements of PI into calculus-based mechanics, the component that can realistically be adopted is the peer-discussion process itself, applied directly to the traditional, mathematically structured problems already used in class.

This distinction is important because most prior research on peer instruction has examined discussion around conceptual multiple-choice questions, whereas much less is known about how peer discussion functions when applied to mathematically structured problems typical of calculus-based mechanics courses.

C. Rationale for examining peer discussion in traditional mechanics problems

PD is a central feature of PI. Previous research has shown that discussion alone—independent of the use of *ConceptTests*—can support conceptual reasoning and promote deeper understanding [18]. However, this evidence comes primarily from discussions of conceptual questions, not from mathematically structured traditional problems. Therefore, it remains unclear whether the benefits of peer discussion extend to traditional mechanics exercises that require mathematical formulation and multistep problem solving. Examining this question is important because such exercises are central to calculus-based mechanics courses and are often difficult to replace with additional conceptual questions under strict curricular constraints.

Because the purpose of this study is to examine the causal impact of PD, we focus here on prior research that used experimental or quasiexperimental designs to compare PD with instructor explanation (IE). Such studies are scarce, and those that exist have been conducted in limited instructional contexts. In physics, experimental evidence on PD is very limited; one of the few studies examined a single problem in electromagnetism [19]. In biology, an often-cited quasiexperimental study compared PD and instructor explanation in genetics [20]. In mathematics, one experimental study examined PD in undergraduate real analysis and reported no significant difference between PD and instructor explanation on final examination performance [21]. This study was conducted in a different disciplinary context and focused solely on long-term exam outcomes rather than immediate problem-level effects. Moreover, most PD tasks used in these prior studies involved conceptual multiple-choice questions rather than mathematically structured problems.

Overall, rigorous experimental comparisons remain rare, and none has examined PD across a range of topics in calculus-based mechanics. Thus, there is a clear need to investigate whether PD is effective when integrated into traditional mechanics exercises and for which types of problems it may be particularly effective or ineffective. Such evidence would provide meaningful guidance for instructors who wish to adopt RBIS within the practical constraints of their courses.

Because many university physics courses face strict curricular and time constraints, it is important to understand how elements of RBIS function when integrated into existing instructional structures rather than implemented in their full canonical form.

D. Purpose of this study

Based on this background, this study examines the educational effects of peer discussion applied to traditional problems in calculus-based introductory mechanics. To investigate the causal impact of PD, this study employs an experimental design based on cluster randomization with a

crossover structure, embedded within an actual university course. To enable a design-based comparison, PD and IE were implemented within comparable problem-solving exercise sessions; procedural details, including how group and order effects were addressed, are provided in Secs. II and III. This design allows PD to be compared directly with IE under controlled instructional conditions.

To our knowledge, no prior experimental study has compared PD and IE across multiple problems in calculus-based mechanics. By analyzing a variety of problem types commonly taught in such courses, this study aims to provide rigorous and practical evidence regarding the conditions under which PD may be more or less effective than IE.

This study addresses the following research questions:

- RQ1: When students first work individually on mathematically structured mechanics problems, how does subsequent peer discussion affect their problem-solving performance compared with subsequent instructor explanation?
- RQ2: For what kinds of problems is peer discussion more effective, or less effective, than instructor explanation?

Section II describes the course and materials, Sec. III explains the research methodology, Sec. IV presents the results, and Sec. V provides the discussion and conclusions.

II. INSTRUCTIONAL CONTEXT AND CLASSROOM PRACTICES

This study was conducted within a compulsory, calculus-based introductory mechanics course for first-year engineering students at a national university in Japan. The course is taught in Japanese and is designed to develop a mathematical and quantitative understanding of classical mechanics. The present study focuses on the first half of this year-long course (an 8-week module), covering kinematics, Newton's laws, oscillatory phenomena, momentum, and work and energy. The module consists of 15 standard 90-min sessions, delivered as two consecutive (back-to-back) classes per week, with a final 90-min end-of-term examination.

A. Participants and instructors

To maintain an optimal class size of 50 students or fewer, multiple parallel classes are offered following an identical syllabus. This study targets two parallel classes within the Faculty of Engineering. Students were randomly assigned to these two classes based on their student identification numbers, resulting in approximately 40 students per class ($N \approx 80$).

Both classes were taught by the authors of this paper. Both are physics education researchers with over 10 years of university teaching experience and extensive prior experience implementing active-learning pedagogies, including PI.

B. General class structure and problem-solving sessions

The course consisted mainly of lecture sessions and problem-solving sessions. In the lecture sessions, the instructor primarily provided explanations while presenting materials using PowerPoint slides. In each 90-min class, approximately four problem-solving sessions were interspersed within the lecture, during which students worked on exercises designed to promote the understanding and retention of key concepts introduced in the lecture.

Peer discussion was introduced during these problem-solving sessions. Following the basic procedure of peer instruction, students first worked individually on a problem for a few minutes (individual round), then discussed it in groups of three to four students (group round), and finally the instructor provided a brief explanation, which together constituted one cycle. As will be described later in the Sec. III A, some of the problems were conducted without peer discussion, and instead, the instructor gave a detailed explanation on the chalkboard.

During the PD sessions, when the problems were multiple-choice, student responses were collected using the learning management system (LMS) before and after the peer discussions. For non-multiple-choice problems, students' confidence levels in their answers were collected. The distribution of responses in the individual round of multiple-choice problems was generally not disclosed to the students. Based on prior experience, when the responses were skewed, regardless of correctness, the majority opinion tended to be regarded as the correct one, which discouraged discussion. This tendency has also been reported by Nielsen *et al.* [22]. However, when the correctness rate was low or the responses were widely dispersed, such information was, as much as possible, provided to the students as feedback prior to peer discussion.

Groups for peer discussion were basically formed at random, and the classes were conducted with assigned seating. A seat change was implemented once during the course. In the first guidance session, the procedure, importance, and significance of peer discussion were explained to motivate students to engage in the discussions. In addition, a short icebreaker activity of about 5 min [23] was conducted before the first class and immediately after the seat change in order to create an environment that encouraged active discussion.

C. Assignments and assessments

The PowerPoint materials used in the course (excluding the exercise problems) were made available in advance on the LMS, and students were encouraged to prepare beforehand. In each class, a preparatory assignment was given, and the first problem-solving activity of the class was based on that assignment. In addition, a homework assignment was given at the end of every class. Homework was

submitted via the LMS, and feedback was provided by teaching assistants under the supervision of the instructor.

At the end of each weekly class, students completed a weekly post-test (hereafter, “post-test”) assessing the day’s content; the post-test was administered for 10–12 min. In addition, a 90-min written final examination was conducted at the end of the semester. Both the post-tests and the final examination primarily consisted of problems similar to those practiced during class sessions. In all cases, the use of notes and textbooks was prohibited, and the assessments were conducted as closed-book examinations.

III. RESEARCH METHODOLOGY

A. Experimental design

As stated in the research questions, the purpose of this study was to quantitatively examine the effects of introducing PD into problem-solving activities in traditional classes.

To clarify the causal effect of PD, we adopted a randomized controlled design. Because students were assigned to two parallel course sections through the university’s course registration process, the study employed a cluster-randomized crossover design at the section level. Specifically, in the classes conducted from June to July 2023, a total of 87 students ($N = 87$) were assigned to Class A ($n = 44$) and Class B ($n = 43$). The instructors of Class A and Class B were S. M. and T. U., respectively. In this design, the instructional conditions (PD and IE) were alternated across exercise topics according to a counterbalanced schedule, so that each section experienced both conditions during the semester.

During the problem-solving sessions, students engaged either in PD or received a detailed IE on the chalkboard, depending on the assigned instructional condition. We focused on the comparison with IE because, in typical classroom practice, adopting PD often requires reallocating time away from instructor explanation. Thus, an important practical question is whether PD is at least as effective as a careful instructor explanation within a comparable time budget. Because both approaches were embedded in in-class problem-solving sessions, individual problem solving was treated as a common component; the comparison focused on the subsequent activity, peer discussion or instructor explanation.

More specifically, PD and IE were conducted as follows. In both conditions, the instructor presented the correct answer at the end of the exercise; the comparison therefore reflects an instructional trade-off under a comparable time budget.

- PD: Brief individual thinking, followed by group discussion and a brief explanation of the solution.
- IE: Extended individual thinking, followed by a detailed, step-by-step worked solution presented by the instructor on the board.

Because the discussion time in PD also included time for individual thinking, the individual round in IE was made slightly longer so that the total time spent on each problem-solving session was approximately the same for PD and IE. Although exact timing was not recorded for every item, both PD and IE typically required about 8–12 min per problem.

In IE sessions, instructors provided primarily explanatory instruction at the board and refrained from interactive questioning except when responding to occasional student inquiries. In PD sessions, instructors followed typical peer instruction practice by circulating during discussions, prompting groups to articulate their reasoning when appropriate and providing minimal hints without disclosing solutions. In both conditions, LMS submissions were collected after the individual round and after the subsequent instructional activity; for non-multiple-choice problems, students reported confidence levels instead of submitting answers. This implementation ensured a clear contrast between PD and IE while keeping the activities close to typical classroom practice.

However, since this study was an intervention in a compulsory course and because we assumed at the planning stage that PD is an effective method, it was not appropriate from an educational ethics perspective to create a class consisting only of PD or only of IE.

Therefore, both classes followed a style that included both PD and IE in the problem-solving sessions. For the target problems, PD and IE were systematically alternated between Class A and Class B, such that one class received PD and the other IE for a given problem, and vice versa for another problem within the same day. The assignment of PD and IE to specific problems was counterbalanced across weeks to minimize potential order effects. Through this rotation of conditions across the two classes, the design functioned as a cluster-randomized crossover structure for the target items.

To ensure consistency across the two classes, the instructors met weekly to review the previous session and coordinate the procedures for the upcoming class. Although the two classes were taught by different instructors, potential instructor effects were mitigated by their similar backgrounds in physics education research and by the counterbalancing inherent in the crossover structure.

Table I shows an example of the class procedure on a given day. In this example, each 90-min class included four exercise problems, and one of them was assigned differently between the two classes, with PD in one class and IE in the other. At the end of each class, a post-test was given on the swapped problems (problems 3 and 7 in this example), and the scores were used to quantitatively evaluate the effects of PD and IE.

In addition to ethical considerations, there are two reasons for focusing on the short-term effects of PD. First, as introduced in Sec. IC, previous studies have

TABLE I. Single-day example of condition assignment in a crossover RCT: each class received PD and IE in opposite order for two problems (PS3, PS7); post-test followed.

		Class A	Class B
First period (90 min)	PS1	PD	PD
	PS2	PD	PD
	PS3	IE	PD
	PS4	IE	IE
Second period (90 min)	PS5	PD	PD
	PS6	PD	PD
	PS7	PD	IE
	PS8	PD	PD
	Post-test (PS3, PS7)		

shown that PD does not have a large impact on final exam results. Second, since factors such as how much each student reviews or studies for the final exam are expected to have a greater influence on the results than the classroom interventions themselves, we considered it important to first clarify how much understanding could be achieved during the class session on the same day. Nevertheless, related problems were also included in the final exam, and medium- to long-term effects were analyzed.

To control for confounding factors other than the comparison between PD and IE, the following measures were taken. As described in the previous section on the course outline, the classes consisted of instructor-led lecture sessions and problem-solving sessions that included about four problems in each class. In the lecture sessions, standardized slides were used, and the instructors followed a fixed script for their explanations.

This research design makes it possible to examine the effects of using PD vs IE in the problem-solving sessions conducted during class. At the same time, it has an important limitation: the classes were not conducted entirely in either PD or IE. Since PD was regularly implemented in both classes, the study does not represent a comparison with a fully traditional lecture composed only of IE. Readers are encouraged to keep this limitation in mind when interpreting the findings.

B. Instruments and data collection

Here, the tests and questionnaires administered in the course as measurement data are summarized.

In the first class, an orientation session was held to explain the objectives and procedures of the course. At that time, three tests were administered as a precourse assessment: (1) the Force Concept Inventory (FCI) [24], (2) the Calculus Concept Inventory (CCI) [25], and (3) the Colorado Learning Attitudes about Science Survey (CLASS) [26]. These tests were primarily used to confirm the equivalence of the two classes that had been randomly assigned. The official Japanese versions were used for the

TABLE II. Dates and brief descriptions of the 16 post-test questions administered during the semester. The full post-tests are provided in Supplemental Material [27].

Date	Item	Description
June 9	Q1	Infinitesimal volume element in three-dimensional polar coordinates
	Q2	Infinitesimal volume element in cylindrical coordinates
June 16	Q3	Motion diagram
	Q4	Differential of polar unit vectors
June 23	Q5	Vector, component, magnitude
	Q6	Select solutions to the differential equation
June 30	Q7	Equation of motion with air resistance
	Q8	Forces acting on a pendulum (polar + xy coordinates)
July 7	Q9	General solution for $d^2x/dt^2 + 4x = 0$
	Q10	General solution of a nonhomogeneous linear differential equation
July 14	Q11	Analysis of damped oscillatory solutions
	Q12	Particular solution of a nonhomogeneous linear differential equation
July 21	Q13	Angular momentum and moment of force in a simple pendulum
	Q14	Line integral of work along a circular path
July 21	Q15	Force from central potential
	Q16	Force and U-contours

FCI and CLASS, whereas the CCI was administered using a Japanese translation prepared by the authors.

In each class, a post-test of about 10–12 min was administered at the end of the session. It was conducted seven times in total, excluding the week of the final exam. The test was in written-response format and focused on the problems for which Class A and Class B engaged in different activities (PD or IE). These tests constitute the main data of this study.

At the end of the semester, a 90-min final exam was administered. The exam was in written-response format and focused mainly on the problems that had been used in the post-tests.

C. Exercise problems and post-tests

The problems used in the problem-solving sessions during class were selected from topics considered particularly important in the course. A variety of problems were included, not only calculation problems but also questions assessing key concepts taught in the lectures. The full list of the problems used in the post-tests is shown in Table II.

Because we were also interested in the effects when problems were implemented in a peer instruction style, some problems were designed as multiple-choice items in the form of concept tests; however, many problems could not be expressed in that format. For such problems, students were asked to report their confidence level in

their answers during the individual round of PD and in the voting after PD.

The complete set of post-test items is provided in Supplemental Material [27]. Grading was conducted by the instructors, with both classes' responses for each item graded by the same instructor using a common rubric. Two graders divided the items and reviewed sample responses together to ensure consistency in scoring.

D. Statistical analysis

Baseline differences between the two classes were examined using independent-samples t tests. For item-level comparisons in the post-tests and final examination, group differences were analyzed using the two-sided Wilcoxon rank-sum test with a significance level of 0.05. Effect sizes (r) were computed for item-level comparisons.

Because 16 item-level comparisons were conducted for the post-test, p values were adjusted using the Holm procedure to control the family-wise error rate [28]. Adjusted p values were compared with an overall alpha level of 0.05.

All statistical analyses were conducted using R version 4.5.1.

E. Ethical review

In this study, an educational experiment was conducted as part of a compulsory course in the Faculty of Engineering at the university. Considering that students were not free to refuse participation in the course itself, we carefully examined the class content and experimental design, as described in Sec. III A, to ensure that (1) no apparent disadvantage would arise between the two classes and (2) the course content specified in the syllabus would be fully guaranteed.

During the first guidance session of the course, the following points were explained to the students, and written consent was obtained.

1. Data from tests and questionnaires would be used for educational research purposes and might be presented at academic conferences or published in journal papers. However, no personally identifiable information would be disclosed.
2. Students had the right to have their own data excluded from the analysis, and such a decision would not result in any advantage or disadvantage.

This study was reviewed and approved by the Ethics Committee of the Graduate School of Humanities and Social Sciences, Hiroshima University, based on the above policies and with appropriate ethical consideration (Approval No.: HR-ES-000931).

IV. RESULTS

A. Group equivalence at baseline

First, the results of the pretests conducted at the beginning of the semester were analyzed. The pretests consisted of the FCI, CCI, and CLASS tests. Figure 1

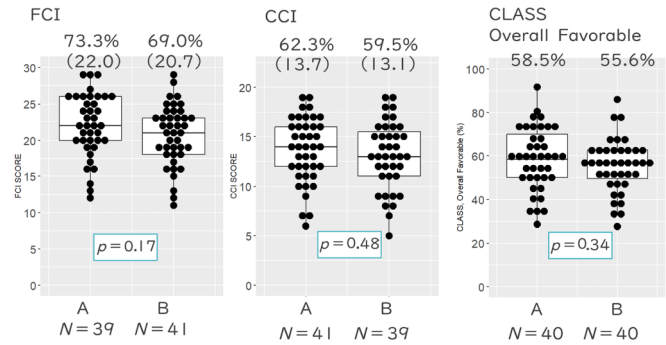


FIG. 1. Pretest results for FCI, CCI, and CLASS (Overall favorable) for Classes A and B at the beginning of the semester. Dots represent individual students, and boxes indicate the interquartile range with median. Mean percentages are shown above each group. Independent-samples t tests revealed no statistically significant differences between the two classes (all $p > 0.05$).

shows the results of these tests. The number of participants slightly varied because a few students were late for the first class session.

The results of the pretests were analyzed using an independent-samples t test. First, regarding the FCI, which measures students' understanding of basic physics concepts, Class A showed a slightly higher mean score than Class B, but the difference was not statistically significant (the significance level was set at 0.05). The mean score of both classes was around 70%, which generally indicates that students are ready to learn Newtonian mechanics. However, the distribution shows that a considerable number of students scored around 50%, suggesting some diversity in their level of understanding.

Next, for the CCI, which measures conceptual understanding of calculus, the mean score was around 60%. Although the validity of this inventory has been discussed, previous research [29] has reported that it is suitable for comparing two groups, and therefore it was used for that purpose in this study. No significant difference was found between the two classes in the CCI scores.

Finally, for the CLASS, which evaluates students' attitudes toward learning physics, the percentage of favorable responses in the "Overall" category was analyzed. Although some differences were observed in the distribution, no statistically significant difference was found, indicating that the two classes did not differ in their learning attitudes.

From these results, it can be concluded that there was no significant difference between the two classes in terms of physics concept understanding, calculus understanding, or learning attitudes and that both classes started from an equivalent level.

B. Results of the post-test

In this section, we report the post-test results, which constitute the primary outcome of this study. The post-test

TABLE III. Post-test results by item and class. Values are mean correct response rates $\pm$ SE(%). For each item, the Condition (A/B) column indicates which instructional method was implemented in Classes A and B. Group differences were tested using the Wilcoxon rank-sum test.

Question	n_A	n_B	A: mean $\pm$ SE(%)	B: mean $\pm$ SE(%)	Condition (A/B)	p	r
Q1	41	41	70.7 $\pm$ 7.2	70.7 $\pm$ 7.2	IE/PD	1.000	0.000
Q2	41	41	53.7 $\pm$ 7.9	26.8 $\pm$ 7.0	IE/PD	0.014*	0.272
Q3	44	43	80.7 $\pm$ 4.7	88.4 $\pm$ 4.4	IE/PD	0.121	-0.166
Q4	44	43	53.4 $\pm$ 4.7	48.8 $\pm$ 5.6	PD/IE	0.542	0.065
Q5	44	42	92.4 $\pm$ 2.1	90.5 $\pm$ 2.8	PD/PD	0.814	0.025
Q6	44	42	23.5 $\pm$ 4.7	26.2 $\pm$ 5.3	PD/IE	0.834	-0.023
Q7	44	42	72.2 $\pm$ 4.5	73.2 $\pm$ 4.7	IE/PD	0.835	-0.022
Q8	43	42	71.3 $\pm$ 4.7	63.1 $\pm$ 4.1	IE/PD	0.080	0.190
Q9	43	42	38.0 $\pm$ 4.6	32.5 $\pm$ 4.9	PD/IE	0.307	0.111
Q10	43	38	93.0 $\pm$ 3.9	73.7 $\pm$ 7.2	IE/PD	0.019*	0.261
Q11	39	38	35.0 $\pm$ 7.3	23.7 $\pm$ 6.3	PD/IE	0.347	0.107
Q12	39	38	35.0 $\pm$ 6.6	30.7 $\pm$ 6.1	IE/PD	0.721	0.041
Q13	42	41	39.3 $\pm$ 5.4	22.0 $\pm$ 3.7	IE/PD	0.033*	0.234
Q14	42	41	65.1 $\pm$ 4.8	59.8 $\pm$ 5.1	PD/IE	0.407	0.091
Q15	42	39	63.3 $\pm$ 5.1	70.3 $\pm$ 5.2	PD/IE	0.277	-0.121
Q16	42	39	82.1 $\pm$ 5.6	53.8 $\pm$ 7.7	IE/PD	0.004**	0.319

Note. p values are unadjusted; asterisks denote unadjusted significance (* $p < 0.05$ and ** $p < 0.01$). Holm correction (16 comparisons) indicated that none of the differences remained statistically significant at $\alpha = 0.05$. Effect size $r = Z/\sqrt{n_A + n_B}$.

was administered at the end of each class to assess students' understanding of that day's lesson.

As shown in Table II, the post-test was administered seven times in total. Each test included two or three questions on the main topic of that day's class, resulting in 16 questions in total. The questions were designed to be similar to the practice problems covered in the class; they addressed the same underlying concepts but differed in surface features.

Except for Q5, the exercises used in the post-test involved different activities between Class A and Class B, incorporating either the PD or IE method. Only Q5 used PD in both classes as a fidelity check, examining whether the same activity implemented in both sections would yield comparable performance.

Table III summarizes the post-test results by item and class as mean correct response rates $\pm$ SE(%). For each item, group differences between Classes A and B were tested using the Wilcoxon rank-sum test, and effect sizes are reported as r . The Condition (A/B) column indicates which instructional method (PD or IE) was implemented in Classes A and B for that item.

Based on the results shown in Fig. 2 and Table III, no significant differences were observed in 12 out of the 16 questions at the unadjusted level. Significant differences were found in four questions: Q2, Q10, Q13, and Q16, all favoring IE over PD.

Because 16 item-level comparisons were conducted, p values were adjusted using the Holm procedure to control the family-wise error rate. After Holm correction, none of the item-level differences remained statistically significant at the 0.05 level.

Although these differences did not survive correction, the consistent directional pattern favoring IE warrants closer examination. In the following, several representative questions—with and without unadjusted significant differences—are described in detail.

1. An infinitesimal volume element in spherical and cylindrical coordinates (Q1 and Q2)

We first discuss the results for Q1 and Q2. In the first week, class guidance and a pretest were conducted during the first period (the first 90 min), so the regular class was held only during the second period (the latter 90 min). Therefore, only one exercise problem was used for the intervention in this session. Accordingly, in the quiz for that week, Q1 was the same as the exercise problem, and Q2 was a similar problem.

In this class, after explaining the Cartesian and polar coordinate representations, we dealt with the method of deriving the area of a circle from the integration of an infinitesimal area element in two-dimensional polar coordinates. Then, as an exercise problem, students were asked to consider the infinitesimal volume element in three-dimensional polar coordinates, while slides such as those shown in Fig. 3 were presented.

In the post-test given at the end of this class, Q1 and Q2 were presented as shown in Figs. 4 and 5. In Q1, the same problem used in the class was asked; that is, students were asked to express the infinitesimal volume element dV in three-dimensional spherical coordinates using r , θ , and φ . Q2 asked students to derive the expression for the infinitesimal volume element in the three-dimensional cylindrical coordinate system.

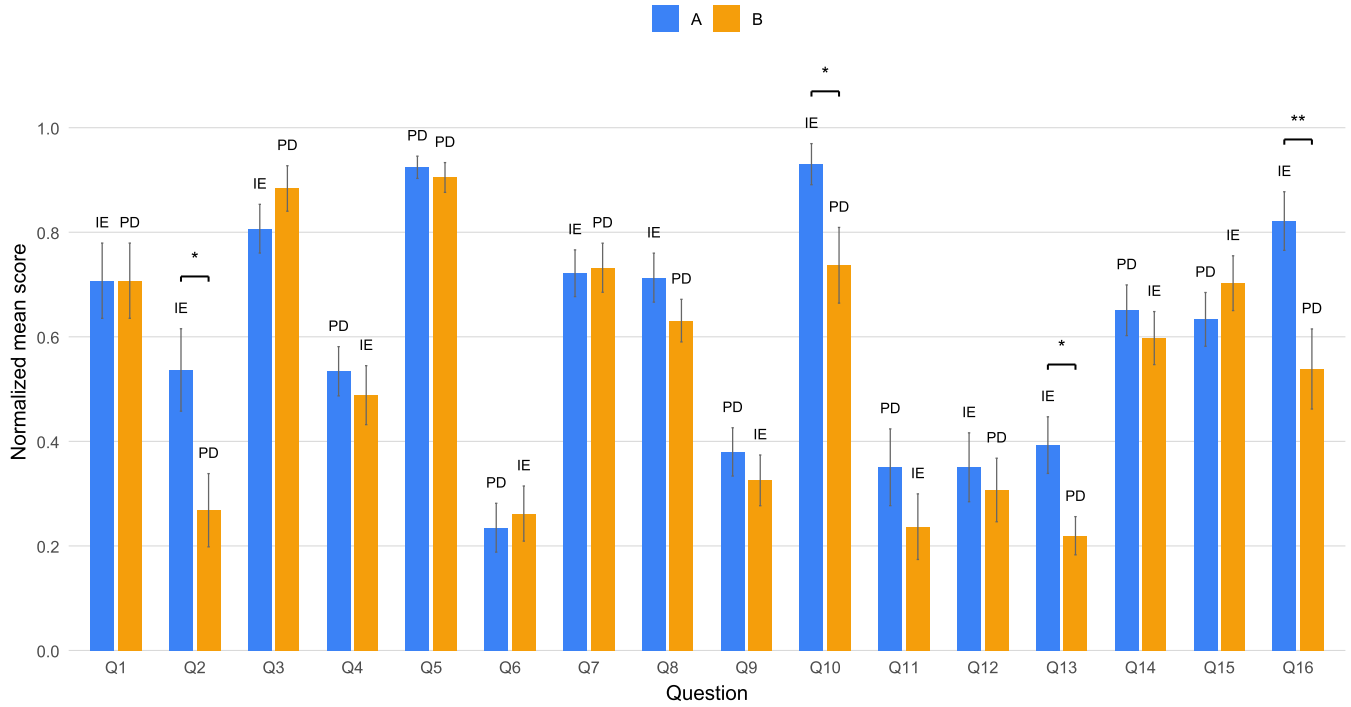


FIG. 2. Mean normalized scores for each question by group (A and B). Error bars represent standard errors. Asterisks indicate significant differences at the unadjusted level based on the Wilcoxon rank-sum test (* $p < 0.05$ and ** $p < 0.01$). After Holm correction for multiple comparisons (16 items), no differences remained statistically significant at $\alpha = 0.05$.

In the class, after the exercise shown in Fig. 3, only a brief remark was made that, in three dimensions, another coordinate system called the cylindrical coordinate system can also be used. No instruction or practice was provided on expressing the infinitesimal volume element in that system. Thus, Q2 served as a transfer item: students who developed a general understanding of the infinitesimal volume element beyond the specific spherical-coordinate form would be more likely to answer Q2 correctly, whereas students who relied primarily on memorizing the expression used in Q1 would be less likely to do so.

2.1.2 Three-dimensional polar coordinates

■ **Question 3**

Find the infinitesimal volume element dV in three-dimensional polar coordinates. (Hint: The dV can be considered approximately rectangular prism.)

- ① $dV = r dr d\theta d\phi$
- ② $dV = r dr \sin \theta d\theta d\phi$
- ③ $dV = r dr \cos \theta d\theta d\phi$
- ④ $dV = r^2 dr \sin \theta d\theta d\phi$
- ⑤ $dV = r^2 dr \cos \theta d\theta d\phi$
- ⑥ no idea

FIG. 3. Slide used during in-class instruction for post-test Q1 (three-dimensional polar coordinates). The original slide was in Japanese; an English translation is shown here. The figure in the question is reproduced from Ref. [30] with permission.

As shown in Table III, post-test performance on Q1 was identical in the IE and PD classes (both 70%), with no evidence of a difference (unadjusted $p = 1.000$). For Q2, post-test performance was higher in the IE class than in the PD class (unadjusted $p = 0.014$). Taken together, these results suggest that the relative effectiveness of PD and IE may depend on item-specific features and how each condition supports students' reasoning.

2. Motion diagram (Q3)

Q3 is a question about motion diagrams, which was covered in the class on vector mathematics. In the class, after explaining the concept and drawing rules of motion diagrams, a motion diagram for an object in uniform circular motion was introduced, and it was explained how centripetal acceleration can be derived from it. Immediately afterward, an exercise like Q3 was given, asking how the motion diagram would look in the case of nonuniform circular motion. However, in the class, the counterclockwise case was used, while in the post-test, as shown in Fig. 6, the clockwise case was presented [31].

A motion diagram is a method for geometrically understanding the relationships among position, velocity, and acceleration through drawing. This problem was selected with the expectation that PD would be effective for students to think about the drawing rules and their meanings through discussion with one another.

Q1: Define the three-dimensional coordinates using r, θ, φ as shown in Figure 2.3 (spherical coordinates). Express the three-dimensional volume element dV in terms of r, θ, φ .

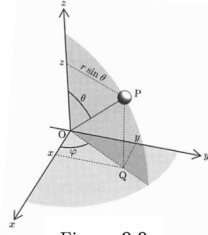


Figure 2.3

FIG. 4. English translation of post-test Q1 (open-ended). The original problem was administered in Japanese; minor adjustments to line breaks and formatting were made for publication. The original version is provided in Supplemental Material [27]. The figure in the question is reproduced from Ref. [30] with permission.

Q2: Define the three-dimensional coordinates using ρ, φ, z as shown in Figure 2.5 (cylindrical coordinates). Express the three-dimensional volume element dV in terms of ρ, φ, z .

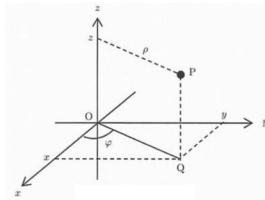


Figure 2.5

FIG. 5. Post-test Q2 (open-ended; see Fig. 4 for translation and formatting details). The figure in the question is reproduced from Ref. [30] with permission.

Q3: Plot the acceleration $a(t)(\Delta t)^2$ for circular motion (clockwise direction; the numbers in the figure indicate the sequence of motion) where the angular velocity decreases at a constant rate, as shown in the figure below. Plot all possible arrows and include not only the final result but also the intermediate steps.

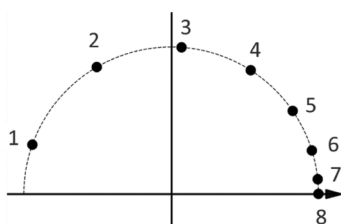


FIG. 6. Post-test Q3 (open-ended; see Fig. 4 for translation and formatting details).

Q10: Consider the nonhomogeneous linear differential equation

$$\ddot{x} + \omega_0^2 x = f_0 \cos \omega t \quad (1)$$

Show that the general solution of this equation can be written as $x(t) = x_p(t) + x_0(t)$, where $x_0(t)$ is the general solution of the corresponding homogeneous linear differential equation

$$\ddot{x} + \omega_0^2 x = 0 \quad (2)$$

and $x_p(t)$ is a particular solution of equation (1).

In other words, $x_0(t)$ and $x_p(t)$ satisfy the following equations, respectively:

$$\begin{aligned} \ddot{x}_0 + \omega_0^2 x_0 &= 0 \\ \ddot{x}_p + \omega_0^2 x_p &= f_0 \cos \omega t \end{aligned}$$

FIG. 7. Post-test Q10 (open-ended; see Fig. 4 for translation and formatting details).

As shown in Table III, although the class that adopted PD had a slightly higher correct answer rate, the difference was not statistically significant. In addition, as shown in the same table, Q3 was the problem for which the class using PD showed a higher correct answer rate and the largest effect size.

3. Solving nonhomogeneous linear differential equations (Q10)

Q10 addressed the technique for solving nonhomogeneous linear differential equations, which is required when solving the equation of motion for forced oscillation. Specifically, it was presented as a written-response problem, as shown in Fig. 7. Although the underlying mechanism of this technique appears complex and difficult at first glance, its principle is actually quite simple, and once understood, the problem can be solved relatively easily.

This problem was selected because it was expected to be a type of task for which PD would be effective. Previous research has shown that when a group works on a mathematical intellectual task and one member proposes a solution that can be readily recognized as correct (i.e., a solution with high demonstrability), the group's performance can exceed that of the best individual in the group [32]. Because Q10 satisfies this condition of high demonstrability, PD was expected to function effectively for this problem.

However, as shown in Table III, the class that implemented IE showed a significantly higher correct answer rate than the class that implemented PD ($p = 0.019$).

For highly demonstrable tasks like this, it is plausible that students can also achieve understanding through IE. On the other hand, because the terminology and concepts involved are abstract, PD—where students can easily ask questions when they encounter difficulties—was expected to work effectively. Therefore, this result was somewhat unexpected.

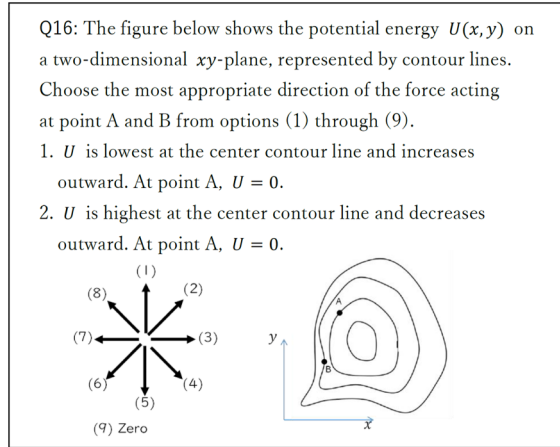


FIG. 8. Post-test Q16 (open-ended; see Fig. 4 for translation and formatting details). Adapted from materials developed by the University of Colorado Boulder PER group [33].

4. Direction of force from potential contour lines (Q16)

The final example, Q16, was the problem that showed the largest difference (i.e., the largest effect size) between IE and PD. This problem was given in the unit on potential energy, where students learned that the force vector acting on an object is given by the partial derivative of the potential energy. Q16, as shown in Fig. 8, also asked students to demonstrate conceptual understanding of this differentiation process. Based on its content, it was expected to be a type of problem for which both PD and IE could be effective.

However, as shown in Table III, the class that implemented IE achieved a significantly higher correct answer rate than the class that implemented PD ($p = 0.004$). In

this unit, students dealt with potential energy in two- and three-dimensional spaces. Because many new concepts were introduced—such as the mathematical operation and interpretation of partial differentiation, and the idea that a force vector is derived from a scalar potential—some students may have struggled to follow the exercises. As a result, PD might not have been carried out effectively in this class.

C. Final examination

In the final exam conducted at the end of the semester, many of the questions were taken from the post-tests. Among the 16 post-test questions, 12 were included in the final exam. The results are shown in Table IV and Fig. 9. As can be seen, there were no significant differences in any of the questions ($p > 0.05$).

This outcome was largely anticipated, since many of the questions had already shown no significant differences immediately after the corresponding lessons. In general, for high-stakes assessments such as final exams, individual factors—such as the extent to which students review the material and prepare for the exam—tend to have a strong influence on performance. As a result, the differences in the effects of classroom interventions are likely to diminish. Furthermore, this finding is consistent with previous research in mathematics education [21], which also reported no significant differences in final exam performance.

V. DISCUSSION AND CONCLUSION

This study compared the learning effects of PD and IE during exercise sessions in a traditional mechanics course. We employed a cluster-randomized crossover design with

TABLE IV. Final examination results by item and class. The examination items correspond to questions previously used in the in-class post-tests. Values are mean correct response rates $\pm$ SE(%). For each item, the Condition (A/B) column indicates the instructional method (PD or IE) originally implemented when that topic was taught in Classes A and B. Group differences were tested using the Wilcoxon rank-sum test.

Item	n_A	n_B	A: mean $\pm$ SE(%)	B: mean $\pm$ SE(%)	Condition (A/B)	p	r
Q2	43	42	84.5 $\pm$ 5.5	86.5 $\pm$ 5.2	IE/PD	0.795	−0.028
Q4	43	42	81.8 $\pm$ 4.8	88.1 $\pm$ 2.7	PD/IE	0.759	−0.033
Q6	43	42	58.9 $\pm$ 5.3	57.9 $\pm$ 5.9	PD/IE	0.936	−0.009
Q7	43	42	58.1 $\pm$ 7.6	57.1 $\pm$ 7.7	IE/PD	0.926	0.010
Q9	43	42	81.8 $\pm$ 3.5	77.0 $\pm$ 4.4	PD/IE	0.647	0.050
Q10	43	42	72.1 $\pm$ 6.9	69.0 $\pm$ 7.2	IE/PD	0.759	0.033
Q11	43	42	32.6 $\pm$ 7.0	40.5 $\pm$ 7.0	PD/IE	0.404	−0.091
Q12	43	42	72.1 $\pm$ 6.9	69.0 $\pm$ 7.2	IE/PD	0.759	0.033
Q13	43	42	49.2 $\pm$ 4.8	41.7 $\pm$ 5.0	IE/PD	0.230	0.130
Q14	43	42	41.9 $\pm$ 5.0	48.4 $\pm$ 4.7	PD/IE	0.289	−0.115
Q15	43	42	63.7 $\pm$ 5.8	61.9 $\pm$ 5.3	PD/IE	0.600	0.057
Q16	43	42	76.7 $\pm$ 6.5	71.4 $\pm$ 7.1	IE/PD	0.578	0.060
Total ^a	43	42	64.5 $\pm$ 2.5	64.1 $\pm$ 2.4	...	0.836	0.022

Note. p values are unadjusted. Effect size $r = Z/\sqrt{n_A + n_B}$.

^aTotal score is expressed as a percentage of the maximum possible score (12).

two course sections, counterbalancing PD and IE across weekly exercise topics. Because the two sections showed no baseline differences on the pretests (FCI, CCI, and CLASS), the subsequent comparisons of short-term outcomes are justified.

A. Revisiting the research questions

RQ1: When students first work individually on mathematically structured mechanics problems, how does subsequent peer discussion affect their problem-solving performance compared with subsequent instructor explanation? In the post-test administered immediately after instruction, 12 of the 16 items showed no statistically significant differences between conditions in unadjusted item-level analyses. The remaining four items favored IE over PD at the unadjusted level (Q2, Q10, Q13, and Q16; Wilcoxon rank-sum test, $p < 0.05$), although none of these differences remained statistically significant after Holm correction for multiple comparisons. Taken together, these results provide no evidence that PD produced higher short-term learning gains than IE under the conditions of this study. As an additional internal check, the similar performance observed for Q5, in which both sections engaged in PD, suggests that baseline section differences were unlikely to account for the observed patterns. More specifically, the unadjusted pattern is consistent with the possibility that IE may be particularly beneficial for items in which an explicit presentation of problem-solving strategies (schema formulation) supports students' reasoning.

RQ2: For what kinds of problems is peer discussion more effective, or less effective, than instructor explanation? Exploratory inspection of item-level patterns suggests that differences between PD and IE may depend on problem characteristics. Because problem features were not systematically manipulated or coded in this study, these interpretations should be regarded as tentative. For example, Q1 (infinitesimal volume element in spherical coordinates) and Q2 (transfer to cylindrical coordinates), which were intended to probe generalization, or transfer, beyond the practiced form, and Q16 (identifying the direction of force from contour lines), which introduced multiple new ideas simultaneously, showed patterns favoring IE. By contrast, for Q3 (motion diagram), which may benefit from visual reasoning and verbalization of problem-solving strategies, the PD group showed slightly higher average scores, although the difference was not statistically significant. Overall, these patterns suggest that the potential advantage of PD may not have been fully realized under the present implementation.

B. Mechanisms underlying the outcomes: Explanatory vs exploratory discussion

One possible reason lies in the implementation of PD. In this study, PD followed a minimal sequence of individual solution—group discussion—short whole-class explanation. Compared with typical peer instruction, the tasks used here less often required students to commit to and publicly contrast alternative lines of reasoning, which may

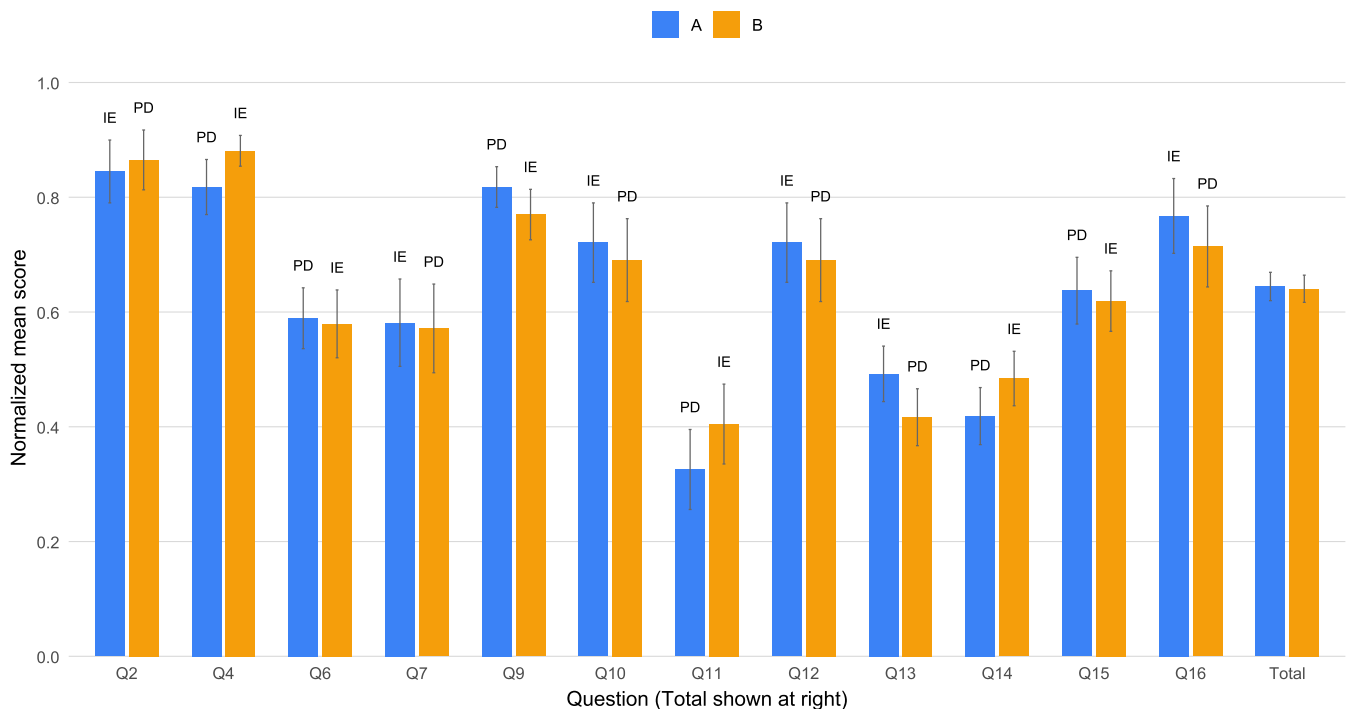


FIG. 9. Mean correct answer rates for each question and for the total score by group (PD and IE). Error bars represent standard errors. No significant differences were observed between groups based on the Wilcoxon rank-sum test (all unadjusted $p > 0.05$).

have limited the depth and structure of peer discussion. Nevertheless, PD activities were relatively active from the beginning of the course. An icebreaking activity was conducted in the first session, and the instructor repeatedly emphasized the importance of discussion, comprehension, and explaining one's reasoning throughout the course. As a result, the class environment was not passive; communication among students was smooth from the early stages. After each PD session, a short instructor explanation was provided to ensure that misunderstandings did not persist. Therefore, the limited effect of PD cannot be attributed to students' reluctance to participate but rather to structural constraints that made it difficult for discussions to reach deeper conceptual or strategic levels.

Additionally, the nature of the problems likely influenced the outcomes. Many topics in traditional mechanics are not designed for conceptual change or correction of misconceptions but rather involve abstract and mathematical ideas requiring students to model and construct their understanding internally. In such problems, students seldom hold clearly conflicting views, making it difficult for discussion to generate cognitive conflict that leads to deeper learning. As a result, discussions tended to become explanatory in nature—where more advanced students explained their reasoning to others—rather than exploratory exchanges of ideas. Under such conditions, explicit explanations and conceptual structuring provided by the instructor (IE) may have been more effective for facilitating understanding.

At least under the instructional conditions of this study, achieving higher learning gains through PD than through IE would require more than the minimal structure used here. It is suggested that further instructional design efforts are necessary to promote deeper and more structured discussions.

C. Effects on high-stakes assessment

In the final examination, which included 12 items from the post-test, no statistically significant differences were observed between the PD and IE groups. Because the final examination was a high-stakes test directly tied to course grades, individual differences in factors such as self-study effort and review behaviors may have had a strong influence, which may have attenuated the observable effects of the in-class intervention. This pattern is consistent with previous findings that the effects of in-class interventions can diminish on high-stakes assessments [21].

D. Limitations and future directions

When interpreting these results, several methodological limitations should be noted. This study adopted a crossover design in which both PD and IE were implemented within each class, rather than a pure comparison between PD-only and IE-only instruction. Therefore, the results do not provide a direct comparison with lecture-only instruction.

Moreover, both classes experienced multiple PD activities, meaning that all students engaged in some degree of active learning; thus, the learning environment differed from a fully unidirectional lecture.

Furthermore, because no direct observation of the PD process (e.g., recording or discourse analysis) was conducted, the quality and characteristics of group interactions could not be systematically evaluated. As a result, it remains unclear whether the observed differences in learning outcomes were due to variation in the quality of PD discussions or to other factors. In addition, the limited number of problems and sample size leave some uncertainty in the estimation of effect sizes. To examine the relationship between PD processes and learning outcomes in greater detail, future research should incorporate discourse and observational data. Such work would contribute to a more precise understanding of the conditions under which PD can function effectively.

E. Conclusion and practical implications

These results highlight the importance of examining how elements of research-based instructional strategies function when implemented under realistic classroom constraints, particularly in mathematically intensive courses.

In conclusion, while peer discussion is widely recognized as a powerful active-learning tool, its integration into mathematically dense and highly structured physics problems does not automatically guarantee superior learning outcomes compared to a well-structured instructor explanation. Our findings suggest that for tasks characterized by high cognitive load—such as those requiring complex mathematical abstractions and novel multi-step reasoning—the explicit scaffolding provided by an instructor may be equally or more effective within limited class time. For practitioners and researchers in calculus-based mechanics, this highlights that simply applying a standard, minimal peer discussion format to mathematically complex exercises is insufficient. To effectively harness the potential of collaborative learning in these contexts, it is necessary to go beyond basic PD and develop more carefully designed instructional interventions to properly support students' deeper reasoning.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grants No. 23K25720, No. 24K00464, and No. 24K21457. We thank the students who participated in this study.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

- [1] M. Prince, Does active learning work? A review of the research, *J. Eng. Educ.* **93**, 223 (2004).
- [2] J. Michael, Where's the evidence that active learning works?, *Adv. Physiol. Educ.* **30**, 159 (2006).
- [3] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. USA* **111**, 8410 (2014).
- [4] Central Council for Education, Towards qualitative transformation of university education for building a new future: Universities fostering lifelong learning and independent thinking (report), 2012, in Japanese.
- [5] 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).
- [6] E. F. Redish and J. R. Burciaga, *Teaching Physics: With the Physics Suite* (John Wiley & Sons, Hoboken, NJ, 2003), Vol. 1.
- [7] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, *Am. J. Phys.* **67**, 755 (1999).
- [8] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, *Am. J. Phys.* **72**, 1147 (2004).
- [9] D. E. Meltzer and R. K. Thornton, Resource letter ALIP-1: Active-learning instruction in physics, *Am. J. Phys.* **80**, 478 (2012).
- [10] C. Henderson and M. H. Dancy, Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations, *Am. J. Phys.* **76**, 79 (2008).
- [11] C. Henderson and M. H. Dancy, Impact of physics education research on the teaching of introductory quantitative physics in the United States, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020107 (2009).
- [12] M. Dancy and C. Henderson, Pedagogical practices and instructional change of physics faculty, *Am. J. Phys.* **78**, 1056 (2010).
- [13] M. Dancy, C. Henderson, and C. Turpen, How faculty learn about and implement research-based instructional strategies: The case of peer instruction, *Phys. Rev. Phys. Educ. Res.* **12**, 010110 (2016).
- [14] C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj, Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process?, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020104 (2012).
- [15] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [16] C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [17] I. D. Beatty, W. J. Gerace, W. J. Leonard, and R. J. Dufresne, Designing effective questions for classroom response system teaching, *Am. J. Phys.* **74**, 31 (2006).
- [18] M. K. Smith, W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Guild, and T. T. Su, Why peer discussion improves student performance on in-class concept questions, *Science* **323**, 122 (2009).
- [19] B. L. Lan, P. M. Lim, and P. W. C. Ho, A modified peer instruction protocol: Peer vs teacher's instruction, *Phys. Teach.* **61**, 290 (2023).
- [20] M. K. Smith, W. B. Wood, K. Krauter, J. K. Knight, and D. Ebert-May, Combining peer discussion with instructor explanation increases student learning from in-class concept questions, *CBE Life Sci. Educ.* **10**, 55 (2011).
- [21] T. Bauer, R. Biehler, and E. Lankeit, Concepttests in undergraduate real analysis: comparing peer discussion and instructional explanation settings, *Int. J. Res. Undergrad. Math. Educ.* **9**, 426 (2023).
- [22] K. L. Nielsen, G. Hansen-Nygård, and J. B. Stav, Investigating peer instruction: How the initial voting session affects students' experiences of group discussion, *Int. Scholarly Res. Not.* **2012**, 290157 (2012).
- [23] Kyoto Sangyo University, Icebreakers you can use on campus!, <https://www.kyoto-su.ac.jp/features/f/action/book.html>, n.d., accessed: 24 September 2025.
- [24] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [25] J. Epstein, Development and validation of the calculus concept inventory, in *Proceedings of the Ninth International Conference on Mathematics Education in a Global Community* (University of North Carolina Charlotte, Charlotte, NC, 2007), Vol. 9, pp. 165–170.
- [26] 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).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/wjhq-n354> for the complete set of post-test items.
- [28] S. Holm, A simple sequentially rejective multiple test procedure, *Scand. J. Stat.* **6**, 65 (1979).
- [29] J. Gleason, S. Bagley, M. Thomas, L. Rice, and D. White, The calculus concept inventory: A psychometric analysis and implications for use, *Int. J. Math. Educ. Sci. Technol.* **50**, 825 (2019).
- [30] M. Inui, N. Hatakenaka, and K. Hoshino, *Kiso Kara Manabu Rikigaku (Fundamentals of Mechanics)* (Baifukan, Tokyo, 2014). in Japanese.
- [31] Since the motion diagram is the same for both clockwise and counterclockwise cases, it was not possible to determine whether some students simply memorized the answer to the problem covered in class and used it to answer this question.
- [32] P. R. Laughlin and A. L. Ellis, Demonstrability and social combination processes on mathematical intellectual tasks, *J. Exp. Soc. Psychol.* **22**, 177 (1986).
- [33] Physics Education Research Group at the University of Colorado Boulder, CU Boulder course materials: Classical mechanics/math methods 1, <https://www.physport.org/curricula/CUBoulder/CM1>, 2014, accessed: 11 March 2026.