Effect of group type on group performance in peer-collaborated two-round physics problem solving

Jiwon Lee [1,*] and Nilüfer Didiş Körhasan [2]

[1]*Department of Physics Education, Korea National University of Education, 28173 Chungbuk, Korea*
[2]*Department of Mathematics and Science Education, Zonguldak Bülent Ecevit University, 67300 Zonguldak, Turkey*

(Received 16 December 2021; accepted 2 August 2022; published 26 August 2022)

Peers play an important role in an individual's learning via collaboration. This study examined three semesters worth of student data collected through collaborative exams in a calculus-based introductory physics course. Participants were asked 142 physics problems over two rounds, and they answered the problems first individually (first round) and then with their groups (second round). Based on the group type determined by the number of correct answers, a majority of correct or incorrect answers, and a variety of answers in the individual round, we examined how group type affects group performance in the group round. A majority of correct and incorrect answers and the variety of answers in the individual round were important for group-round performance. The percentage of finding the correct answer in the group round was directly proportional to the number of students with correct responses in the individual round. Furthermore, an increase in the number of correct answers in the individual round increased the group performance when similar conditions were seen in the majority of correct and incorrect answers and the variety of answers in the previous round.

DOI: 10.1103/PhysRevPhysEducRes.18.020112

I. INTRODUCTION

Social constructivist theory states that learning depends on interactions with others, thus explaining learning as an interactive process that considers individuals as part of a social group. With this focus on interaction, knowledge is defined as a human product [1] that is constructed in a social environment [2]. As such, social interaction primarily helps to construct new meanings [2].

Classrooms are important social environments, providing opportunities for communication and interaction [2,3]. Furthermore, working with peers may contribute towards individuals' learning in several ways, such as problem solving, activation and construction of knowledge, and knowledge exchange and revision. Numerous studies have indicated the importance of peer collaboration, even when the members of a group do not know the correct answer individually [4–6]. In this study, we focused on peer collaboration in two-round collaborative exams, whereby students answer physics problem sets individually, discuss it with their group members, and resubmit group answers on the same problem set. With the identification of group types that are determined by considering the number of correct answers, a majority of correct or incorrect answers, and the variety of compositions in the individual round, we examined how group performance, the percentage of groups that find the correct answer after discussion, was affected by group types. The research questions for investigation are as follows:

How is the group performance dependent on the

- *number of correct answers in the individual rounds of the group members?*
- *majority of correct or incorrect answers in the individual rounds of the group members?*
- *variety of answers in the individual round of the group members?*

The answers to these research questions could help design effective and improved collaborative environments for students' physics education by minimizing the disadvantages related to group characteristics.

*jiwonlee@knue.ac.kr

II. LITERATURE REVIEW

A. Why and how is peer collaboration important for physics learning?

Johnson and Johnson explained three basic ways students interact when they learn science. They are competition, where some students won and some students lost; individual, where each student had their own goal; and working together to attain a shared goal, where all group members shared effort and reward [7]. The same researchers also showed that students learned more, enjoyed the subject matter more, and had higher levels of self-esteem when they worked together by examining much research in different subject matters and age levels [8]. During a collaborative interaction, individuals may interpret situations, find solutions to problems together, and may coordinate their understanding [9]. In classrooms, peer collaboration may facilitate individuals' learning, as peers reduce the cognitive load during collaboration and maintain individuals' attention during peer interaction [5]. Additionally, peers provide each other with a potential development zone (zone of proximal development [2]) in the collaborative process, which could not be reached through an individual's actual performance. Peers also provide social persuasion via verbal or nonverbal judgments, which is important for self-efficacy. Self-efficacy refers to individuals' ideas about their ability to complete specific tasks and is a strong predictor for learning [10–13]. Language is another important element in cognitive development and individuals' learning [2] as it allows the exchange of knowledge and deepens understanding [14–15]. Owing to the use of informal language instead of formal [16], students might explain concepts to each other more effectively than their instructors [17]. Since they are aware of each other's difficulties, they can better transmit the instructor's expectations [5].

In physics education, some researchers demonstrated how students presented four different types of discussions while collaborating on working tutorials [18]. This provides a foundation for physics-related discussions, such as asking and answering questions from peers and instructing others. There are also discussions related to practical issues and meta-level discussions, including metacognitive and metaconceptual issues, such as reflection, disagreement, confirmation, explanation, and elaboration of the discussed topics. Another study indicates three dimensions of collaboration: social, discursive, and disciplinary content [19]. Previous studies in physics education have also pointed out how peer collaboration could be important for learning, such as for the activation, co-construction, revision, and enhancement of physics knowledge.

Knowing how and when knowledge is activated and used is important for good knowledge organization [20]. Knowledge can be activated by situational and contextual factors in an environment [21]. Peers may trigger the recall of relevant concepts during interaction [5]. When we consider the activation of cognitive resources, students' wrong answers can be considered as the application of inappropriate knowledge, which is useful in another context or for solving a given problem. Wood *et al.* examined first-year undergraduate physics students' peer discussions in the implementation of Peer Instruction [21]. Approximately 200 students who took a calculus-based physics course voluntarily formed small groups. Seven groups (with a total of 53 students) were examined during three lectures composed of seven Peer Instruction sessions. The researchers electronically recorded students' discussions (written notes and dialogue). Additionally, students' answers after the discussions were recorded by means of a digital voting system. The researchers observed that students activated each other's cognitive resources during peer discussion in three ways: by activating knowledge elements, which are declarative and procedural knowledge held in long term memory; knowledge structures, which are the linking patterns of association between knowledge elements; and control structures, which function to determine when these elements and structures are activated. In almost half of the transcribed dialogues, at least one example of the activation of resources can be detected for each question.

By interacting with each other, individuals can organize their knowledge through the revision of previous knowledge or by constructing new ones. Knowledge convergence is among the theories pointing to the social nature of knowledge construction; it is explained as the increase of common knowledge of all collaborating partners in a small group after collaboration [9]. From this perspective, Singh focused on university students' co-construction of physics knowledge [5]. She described the co-construction of knowledge in terms of the performances of peers in a two round (individual and peer) conceptual test. As a result of the co-construction of physics knowledge, the group arrived at the correct conclusion even though nobody in the group knew the correct answer. In the researcher's experiment, in a calculus-based introductory physics course, the students were allowed to select their peers and 37 pairs are accordingly formed for the experimental group. They partook in a peer discussion on the "Conceptual Survey of Electricity and Magnetism" test questions [22]. The experiment findings indicated significantly higher scores for peer collaboration than individual performance. The author identified co-construction by noting that students with incorrect answers could find the correct answer by collaborating in groups, even if the incorrect answers were identical. Additionally, the study revealed how students valued peer interactions. Singh explained that co-construction helps in facilitating articulation and enabling students to make sense of their own thoughts in peer discussions [5].

Bungum *et al.* examined how peer discussions enhanced upper secondary school students' understandings of quantum physics concepts [14]. In total, the researchers analyzed 96 peer discussions between two or three students about wave-particle duality and Schrödinger's cat, topics that are abstract and raise epistemological and ontological questions. The students also provided independent statements and confirmatory explanations. The researchers found 70% of the discussions to be productive, with cumulative and exploratory explanations that deepened the understanding of abstract concepts in quantum theory through knowledge exchange. As such, the researchers concluded that the use of small group discussions enhanced students' understandings of abstract concepts. Therefore, these are worthwhile in physics teaching, supporting the role of language in learning, as described by Vygotsky [2].

Kalman *et al.*'s experimental study indicated how collaboration promoted conceptual change in an introductory calculus-based mechanics course [23]. The researchers took a collaborative group approach for an experimental group and a professor-centered or traditional approach for a control group. Three or four students were assigned to each group. Conceptual conflict was set and groups debated on the issues. Post-test scores indicated a statistically significant gain for those engaging in collaborative work while solving physics questions, rather than undertaking individual work.

Collaboration improves students' qualitative problem-solving abilities as well [24–25]. Mason and Singh examined how problem solving with peers could be beneficial for students' physics problem solving [26]. The researchers hold an experiment on an algebra-based introductory physics course with 200 students. While the traditional recitation hours—as upheld for a control group—included teaching assistants answering students' homework questions and implementing a quiz (107 students). In the experimental group, recitations were designed with peer reflection, providing guidance for students via reflective problem solving with their peers (97 students). In the peer-reflection group, students formed nine groups and worked together to discuss the best solution with their peers. Although the instructor emphasized effective problem-solving strategies during the semester to the control and experimental groups, Mason and Singh identified how the peer-reflection group solved more problems with diagrams than the traditional recitation group; this difference was statistically significant [26]. When the students' problem solutions were examined, the peer-reflection group was noted to draw more diagrams than the control group. Additionally, the correlation between the number of problems solved with diagrams drawn and the final exam score was found to be higher for the peer-reflection group than the traditional one.

Heller *et al.* also designed an experiment to investigate the effects of solving context-rich physics problems with peers [15]. In an algebra-based physics course, students were taught a problem-solving strategy with five steps. Subsequently, students worked with their peers and practice the problem-solving strategy. The students went on to solve a test in two parts—one with the group and the other individually. They took it during recitation hours with their peers, and then individually the following day during lecture hours. Hypothesis testing of the group and individual problem-solving scores of the best problem-solver in each group indicated that group solutions were statistically significant with better scores versus the individual solutions of the best problem-solvers. Heller *et al.*'s results conclude that the occurrence of collaboration gave better solutions to solve the physics problems [15].

Harlow *et al.* examined the role of the diversity of group members in terms of ability, the number of students in the groups, and the number of male students in the group, and the effects of them on the groups' performance during collaboration; they demonstrated that none of these elements significantly impacted student learning [27].

B. Do students learn physics via two-round collaborative exams?

In comparison to the learning process, testing has long been thought to be an individual one. However, two-round collaborative exams have been devised and widely applied as a learning strategy to gain immediate feedback from peers [28]. Many studies demonstrated improved cognitive gains in the group round, corresponding to group performance, when applying two-round collaborative exams [29,30]. These gains are considered the result of three mediators: cognitive processes such as remembering information better, having good discussions and increased ability to think about information, interpersonal interactions such as working successfully with others, and reduced test anxiety between the individual and collaborative portions on the exam [31].

Some of the studies indicated that the content knowledge learned via two-round collaborative exams is retained for a long time [32,33]. However, some of the researchers argued that two-round collaborative exams improved performances but were not linked to retaining knowledge [34]. This debate about content retention raised the question of whether the performance improved in two-round collaborative exams because of collaborative learning or because students tend to cheat. Jang *et al.* examined the performance of engineering and pre-medical students in a calculus-based mechanics course [4]. In two rounds, 67 students solved open-ended physics problems, first individually, and then with group collaboration. In total, 14 teams responded to 31 questions; 26% of the 434 team responses came from the groups with students who gave entirely wrong answers in the individual round. After the first team attempt, 25%, after the second attempt, an extra 20%, and finally, after the third attempt, another 7% of them obtained the correct answer for each stage. In the end, after three attempts, 52% of the teams with all-wrong answers in the individual round obtained the correct answer. This implied that teams could generate the correct answer even when there was no correct answer at the beginning of the group round. Furthermore, Jang *et al.* demonstrated how best-in-team students gained from collaboration, with statistically significant better group-round scores. On observing an increase in the number of correct answers after collaboration in both groups with strong members and members who did not provide the correct answer, the researchers concluded that collaboration rather than the propagation of correct answers was responsible [4].

While many studies on two-round collaborative exams for problem solving report the cognitive benefits of collaboration, the nature of the answers given in the individual round are not examined. This study investigates peer collaboration in two-round collaborative exams within the context of an undergraduate physics course. It explains how group performance changes owing to group type, which is determined by the number of correct answers, a majority of correct or incorrect answers, and a variety of answers in the individual round. As it focuses on the above-mentioned factors, this study differs from previous research.

III. METHODOLOGY

We define collaboration as whether students in a peer group arrive at either the correct or wrong answer after peer discussion with a single shared goal "to answer the question" in the two-round collaborative exam. A two-round collaborative exam is an assessment approach where individuals work together on examinations [31]. In the individual round, participants solve the set of questions alone and submit individual answers. In the group round, group members discuss each question and submit one group answer. The rate at which a group chooses the correct answer after a discussion is called group performance. Group types are classified according to the individual-round responses of group members.

This study examines group performance through the determination of group types based on the number of correct individual answers, a majority of correct or incorrect individual answers, and a variety of individual answers.

A. Research setting and data collection

Data were collected over three semesters from different cohorts of students taking a calculus-based introductory physics course at Harvard University. In addition to the implementation of Peer Instruction (PI) via modern platforms, the use of tutorials, estimation activities, experimental design activities, problem-set reflection, and collaborative exams focusing on solving physics problems in two rounds were other important components of the course. In each semester, there were around 60 students taking the course. Their majors were engineering, pre-medicine, humanities, etc.

1. Two-round collaborative exams

For a course assessment, students took two-round collaborative exams [28]. In this study, two-round collaborative exams are called "readiness assurance activities (RAAs)." Similar to the implementation of PI, RAA questions are released by the instructor via a web-based electronic response system. The process of RAAs is shown in Fig. 1.

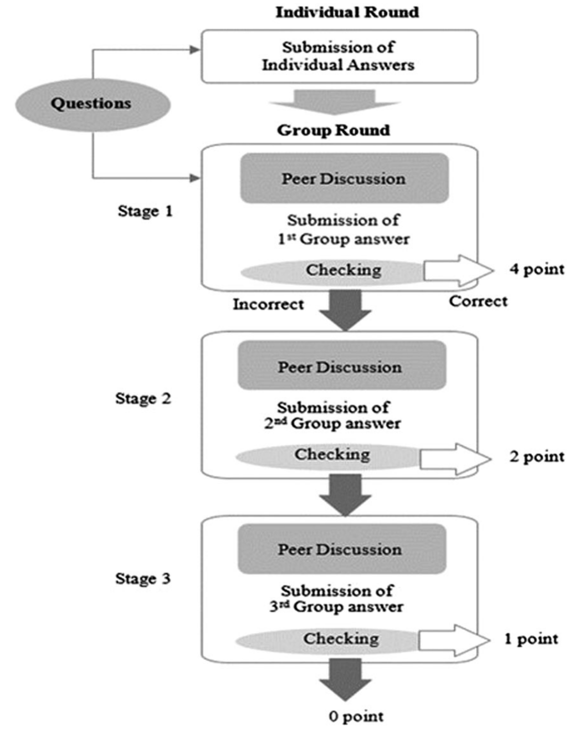

FIG. 1. Process of two-round collaborative exams, RAA. 4–5 students who submitted individual-round responses form a group, conduct group discussion, and then submit group-round responses. The group that fails to submit the correct answer in the first attempt of the group round gets two additional chances.

The types of questions given to students in RAA are not just numerical but include both analytical and conceptual types. In individual rounds, students submit their own answers to given questions. Students then form groups, discuss the same questions with group members, and submit the group's responses to the system. Only one group answer is submitted for each question. After each group response, instant feedback is given on the students' answer. The system checks whether the group answer is correct or not. If students answer the question correctly in the first step of the group round, the group gets four points (stage 1); if they do not, the students are asked to answer the question again after peer discussion. If the group answer is correct in the second step of the group round, the group gets two points (stage 2); if not, the students are given a last (third) chance to discuss it and answer again. If the group answer to the question is correct, the group gets one point (stage 3); if not, the group cannot receive any partial credit (fail). Thus, the group round may extend up to three stages if students cannot find the correct answer after discussion. If you fail to submit the correct answer three times, the system will give the correct answer.

Although PI and RAA activities involve peer discussion in two rounds, the difference between PI and RAA is how questions are posed. In PI, students answer a question (Q1) individually, then share their opinions and ideas in a discussion, and finally, revise their answers (for Q1) individually. The same goes for all other questions. However, in RAAs, the students first solve all the problems on their own in the individual round (Q1, Q2, Q3, etc., before discussing their answers in the group round). Subsequently, they discuss and choose a group answer for each question. This process is similar in the group round for all the questions.

2. Grouping

RAA groups are usually composed of four or five students as with PI. The formation of the groups is determined by the instructor. In a semester, five RAAs are set for students. Table I presents the number of peer groups and RAA questions over the three semesters in our study.

Even in a group composed of the same members, the individual-round response pattern of group members varies according to each problem, so the group type may be different. According to Table I, over three semesters, we obtained a total of 2057 individual and group responses, and except when no group answer was submitted, 2047 instances were analyzed.

B. Data analysis

This study investigates the difference in group performance according to the group type. Group types can be divided according to how group members respond in individual rounds. Additionally, the group performance is determined depending on whether the group submits the correct answer after discussion.

1. Group types

Group types classify the individual-round responses of group members into three dimensions: number of correct individual answers in the group, majority of correct or incorrect answers, and variety of individual answers in the group.

Number of correct answers in the individual round.— For each group type, the number of correct answers in the individual round was considered, no matter if the group-round answers were correct or incorrect. There were mainly four groups based on the number of correct answers in the individual round:

TABLE I. Number of peer groups and RAA questions. G: Group, Q: Questions. Each RAA includes 7 to 11 problems on related physics content. There is a minimum of 14 and a maximum of 16 groups of students answering the questions in each RAA set for each semester.

Semester	RAA1	RAA2	RAA3	RAA4	RAA 5
Spring 2015	14G*11Q	14G*11Q	14G*11Q	15G*9Q	15G*10Q
Fall 2015	14G*9Q	15G*7Q	15G*8Q	16G*8Q	15G*8Q
Spring 2016	14G*10Q	15G*11Q	14G*11Q	14G*8Q	14G*10Q

(1) *Type A*, where all group members provided the correct answers to questions in the individual round.
(2) *Type B*, where more than half of the group provided the correct answer or equal number of correct and incorrect answers in the individual round.
(3) *Type C*, where the correct answers were provided by less than half in the individual round.
(4) *Type D*, no one in the group provides the correct answer in the individual round. That is all incorrect answers case.

The number of correct answers is important to discuss the group performance according to group type considering both "majority" and "variety."

Majority of correct or specific incorrect answers in the individual round.—This is about which type of answer (correct or one type of incorrect) is dominant. It is possible to divide groups based on whether the number of correct answers exceeds the number of same incorrect answers, or vice versa, or whether the number of correct answers and same types of incorrect answers are equal. This classification is different from the above classification by number of correct answers. For instance, if the number of correct answers in a group is less than half but there are multiple types of incorrect responses, the correct answer constitutes the majority in that group.

(1) *C*, majority is the correct answer, where the correct answers provided in an individual round are more than the number of incorrect answers.
(2) *T*, tie between the correct and specific incorrect answer, where the correct and same incorrect answers were provided by an equal number of members in the individual round.
(3) *I*, majority is a specific incorrect answer, where the number of incorrect answers in a given round exceeds the number of correct answers.
(4) *N*, nonmajority, no individual answers have the majority because all individual answers are different.

Variety of answers in the individual round.—We examined the extent to which the answers were varied in the individual round. Within a group, individual-round answers might be in different variations, such as all the same, all different, or all partially different. Thus, variety refers to the type of individual-round responses, regardless of whether they are correct or incorrect answers. For example, for a given problem, there is only one correct answer, but there might be only type of one incorrect answer, or several incorrect answers that are different from each other. We analyzed the variation of these answers into three categories:

(1) *S*, where all answers are the same,
(2) *P*, where some answers are different,
(3) *D*, where all answers are different.

As presented in Table II, based on these three dimensions, different group types might be displayed just before the group round, which have probable influence on group performance during discussions.

TABLE II. Group types based on individual-round performance. In the symbolic form of the individual-round answers, X is same incorrect, x is different incorrect, and O is the correct answer. Different from the group type due to the number of correct answers, group type due to a majority and variety consider "same" and "different" natures in the incorrect answers.

Number of correct answers	Majority	Variety	Symbolic form of individual-round answers		
Type A	C *(Type A-C)*	S *(Type A-S)*	OOOOO		OOOO
Type B	C *(Type B-C)*	P *(Type B-P)*	OOOOX		OOOXX
			OOOxx		OOOx
				OOxx	
	T *(Type B-T)*			OOXX	
Type C	C *(Type C-C)*	P *(Type C-P)*		OOxxx	
	T *(Type C-T)*			OOXXx	
	I *(Type C-I)*		OOXXX		OXXXX
			OXXXx		OXXxx
				OXXx	
	N *(Type C-N)*	D *(Type C-D)*	Oxxxx		Oxxx
Type D	I *(Type D-I)*	S *(Type D-S)*	XXXXX		XXXX
		P *(Type D-P)*	XXxxx		XXXxx
			XXXXx		XXxx
				XXXx	
	N *(Type D-N)*	D *(Type D-D)*	Xxxxx		Xxxx

2. Group performance

This study deals with only the group performance after the first discussion. Group performance is calculated as the percentage of groups that find the correct group answer after first discussion among instances by group type. The group performance could be divided into two phases: performance after the first discussion and performance after the second and third discussions. For the first discussion, group members consider only their individual responses and their conversation, not additional information from the system. However, in the second and third discussions, a group discussion is conducted after receiving the feedback that their submitted answer is incorrect. Therefore, the submission rate of correct group answers after the first discussion was used to analyze the group performance as a difference in discussion results due to group types, without additional information from the system.

IV. RESULTS

Group performance was discussed in three parts based on the number of correct answers, majority of correct or incorrect answers, and variety of answers in the individual round.

A. How does the group performance differ according to the number of individual-round correct answers in the group?

The group performance, the percentage of groups that find the correct group answer by group type, is higher depending on the number of correct respondents in the preceding individual round. Moreover, assuming that individual responses within a group have the same weight, the observed rate of group performance of types B, C, and D, except for type A, exceeded the expected rate range. Table III shows how group performance varies with the number of correct answers in individual round.

In the type A group, where all answers were correct in the individual round, assuming that each individual answer within the group had the same weight, the expected percentage to choose correct group answer was 100%. In fact, 97.59% of the groups succeeded in finding the correct answer in the first stage of the group round.

For type B, the percentage of correct answers in the first stage of the group round was 94.57%. As equal or more than half the answers in the individual round were correct answer, the probability of the correct answer being found in the group round was more than 50% and less than 80%. The group performance of type B exceeds the expected range, assuming that individual responses have the same weight.

TABLE III. Group performance according to the number of individual-round correct answers in the group. OR: Observed rate of group performance. ER(W): Expected rate based on equal weight of individual answers.

Group type	Number of groups (%)	OR %	ER(W) %
Type A	166 (8.11)	97.59	100
Type B	700 (34.20)	94.57	50–80
Type C	743 (36.30)	76.58	20–40
Type D	438 (21.40)	30.82	0

Type C refers to groups where the correct answers were provided by less than half of the members in the individual round. The percentage of correct answers of the group round was initially at 76.58%, which is lower than that of types A and B's group performance. However, this observed rate was higher than the 20%–40% range of expected rate based on equal weighing of individual responses.

Type D consisted of respondents who did not report a correct answer in an individual round. Therefore, the expected rate is 0% in this group type. However, 30.82% of type D group found the correct group answer through first group discussion outside the range of responses provided by individual group members.

B. How does the group performance differ according to the majority of individual-round correct or incorrect answers in the group?

1. The group performance according to the majority within group types

When individual-round responses of group members have equal weights, the observed rate of group performance far exceeds the range of expected rates. To determine the reason, the group's correct and incorrect answers were divided into four categories: majority with correct answers (C), majority with incorrect answers (I), tie between correct and incorrect answers (T), and nonmajority of individual answers (N). The reason for this division is that there is only one type of correct answer, but there can be one or multiple incorrect answers. Therefore, even if there are few correct answers in the group, if there are various incorrect answers, the correct answer can be majority (see symbolic forms Table II). Figure 2 shows how the group types are divided and how the group performance varied with the majority of individual answers.

Table IV presents the results of chi-square analyses for the majority. Type A is not divided because all answers are the same correct answers (that is, Type A-C). Type B is divided into type B-C and type B-T. The group performance of type B-C was 95.85%, which is statistically higher than that of type B-T, 83.78%. Type C is divided into type C-C, type C-T, type C-I, and type C-N. The group performance was in the order of type C-C (87.40%), type C-N (80.71%), type C-T (74.07%), and type C-I (65.63%), respectively, and the group performances are statistically different. Type D is divided into type D-I and type D-N. Furthermore, in type D, the group performance of type D-I which is the majority with incorrect answers is 23.38%, whereas the group performance of type D-N, in which none of the answers constitute a majority is 38.78% statistically higher.

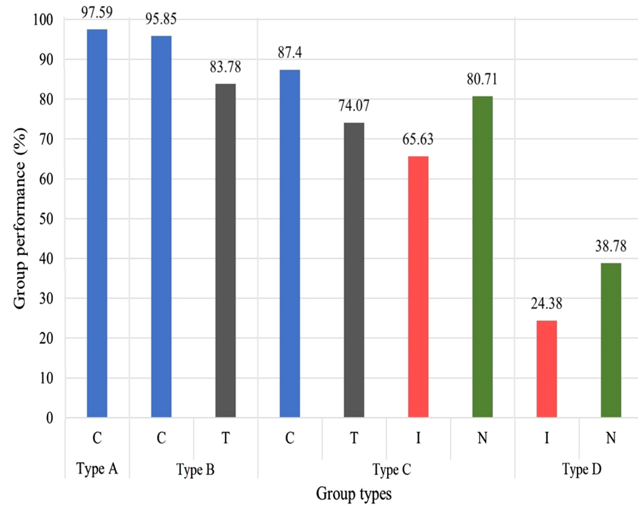

FIG. 2. Group performance according to a majority of correct or incorrect answers of group members' individual-round answers.

In summary, the group performance showed the highest majority with correct answers, followed by a nonmajority of individual answers, a tie between the correct and incorrect answer, and a majority with incorrect answers. In addition, as shown in Table IV, even though the number of correct answers in the group is the same, the group performance was statistically different depending on the majority.

2. The group performance of a group type with the same majority

When correct individual answers are the majority in the group, do groups make decisions by majority vote?—If

TABLE IV. Group performance according to the majority of individual-round answers in the group. α is 0.05.

Group type		Number of groups (%)	Group performance %	χ^2
Type B	C	626 (89.43)	95.85	18.757, $df = 1$, $p < 0.001$
	T	74 (10.57)	83.78	
Type C	C	127 (17.09)	87.40	26.520, $df = 3$, $p < 0.001$
	T	81 (10.90)	74.07	
	I	224 (30.15)	65.63	
	N	311 (41.86)	80.71	
Type D	I	242 (55.25)	24.38	10.525, $df = 1$, $p = 0.001$
	N	196 (44.75)	38.78	

groups decide by majority vote, the expected rate of group performance is 100%. The observed rate, as shown by the blue bar in Fig. 2, did not reach 100% and decreased as the number of correct answers in the group decreased. Comparing the group performances of type A-C, type B-C, and type C-C, type A is the highest with 97.59%, followed by type B-C with 95.85%, and type C-C has the lowest with 87.40%. There was no statistical difference between type A and type B-C ($\chi^2 = 1.095,\ df = 1,\ p = 0.295$), but there was a statistical difference between type B-C and type C-C ($\chi^2 = 14.298,\ df = 1,\ p < 0.001$). That means when the number of correct answers is dominant to the same incorrect answers or each different incorrect answer, group performance is better for the groups that include equal or more correct answers (types A and B) than the groups with less than half correct answers (type C) in overall.

When the correct answers and the coinciding incorrect answers are tied, are group performances tied?—The group performance can be expected to be 50% if the correct and incorrect answers are considered as equal weights. However, as shown by the black bars in Fig. 2, the observed rates are higher than the expected ratio range in both groups. The group performance of type B-T is 83.78%, and type C-T is 74.07%. There is no statistical difference between the two groups' performances ($\chi^2 = 2.176,\ df = 1,\ p = 0.140$). This means that when the number of correct answers and the same incorrect answers are equal (tie), group performance is similar for the groups that include equal or more correct answers (type B) and the groups with less than half correct answers (type C) overall.

When specific incorrect individual answers form a majority in the group, do groups make decisions by majority vote?—If a group answer is decided by a majority vote, the expected rate is 0% because there is a certain incorrect answer that prevails. When choosing individual answers with equal weight, the expected rates are 20%–25% for type C-I and 0% for type D-I. However, as shown by the red bars in Fig. 2, the observed rates are higher than the expected ratio range in both groups. The group performance of type C-I is 65.63%, and type D-I is 24.83%. The difference is that type C-I includes correct answer(s), whereas type D-I has no correct answer. There is a statistical difference between the two groups' performances ($\chi^2 = 15.549$, df $= 1,\ p < 0.001$). This implies that when the majority is the same incorrect answer, group performance is better for the groups including a correct answer (type C) than the groups without a correct answer in the individual round (type D) overall.

C. How does group performance differ based on the variety of answers in the individual round of the group?

It is confirmed that the group performance of the group types with a high proportion of correct answers or majority in correct answers was relatively high. However, when group members' individual-round responses differ, group performance is high despite the small number of correct answers or even the absence of a correct response. Therefore, it is necessary to analyze how does the group performance differ according to the variety of individual answers.

1. The group performance based on the variety within group types

We analyzed how the group performance varies according to the variety of answers within the group types that are divided according to the number of correct answers in the group. Combinations due to the variety of individual answers were divided into three types: all the same, partially different, and all different. Figure 3 shows how group types are divided and how group performance varied with the variety of individual answers. In addition, Table V presents the results of chi-square analyses showing that the performance differs owing to variety.

As shown in Table V, even if the number of correct answers in the group is the same, the difference in the group performance was statistically significant depending on the variety of answers in the group.

By considering group types (see Table II) due to a variety of answers, Type A-S is all the same, and Type B-P is partially different. These groups are not divided into subgroups by variety.

Type C, where the correct answers were provided by less than half in the individual round, is divided into type C-P and type C-D. Group performance of type C-D, where all answers are different, including only one correct answer (80.71%), is higher than type C-P which has some answers that are statistically different (68.66%).

Type D, with no correct answer in the individual rounds, is divided into three subcategories according to the diversity of incorrect answers: Type D-S, type D-P, type

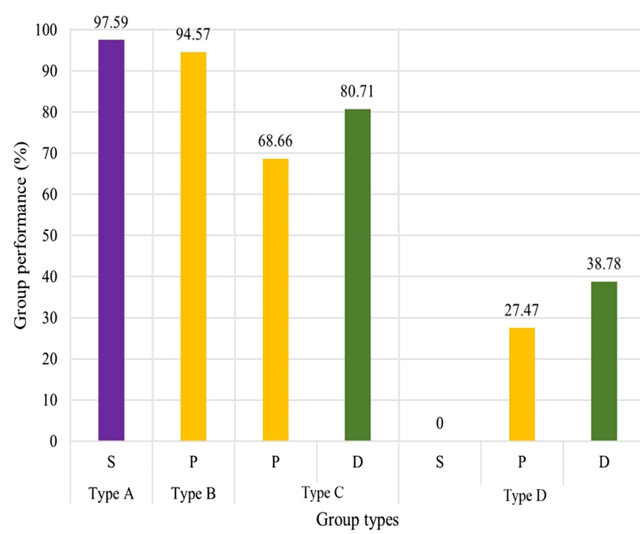

FIG. 3. Group performance according to a variety of group members' individual-round answers.

TABLE V. Group performance according to the variety of individual-round answers in the group. Type C-D is the same as type C-N, also, type D-D is the same as type D-N. α is 0.05.

Group type		Number of groups (%)	Group performance %	χ^2
Type C	P	432 (58.14)	68.66	5.077 $df = 1$, $p = 0.024$
	D	311 (41.86)	80.71	
Type D	S	18 (4.11)	0.00	15.946 $df = 2$, $p < 0.001$
	P	224 (51.14)	27.47	
	D	196 (44.75)	38.78	

D-D. The highest group performance is type D-D, where all incorrect answers differ (38.78%), then type D-P, the group in which there were some same incorrect answers and some different incorrect answers (27.47%), and finally type D-S, the group which had all the same incorrect answers (0%).

In other words, when there are no correct individual answers and when there is only one correct answer in the group, the rate of finding the correct group answer is higher as the incorrect answers are more diverse, and, consequently, the group performance is higher.

2. The group performance of a group type with the same variety

When individual-round answers are all the same, will there be consensus in the group round?—Consensus occurred in most cases when all individual-round answers within a group were the same. Cases with the same individual answers were divided into two categories: coincident with the correct answer (type A-S), and coincident with an incorrect answer (type D-S). As shown by the purple bars in Fig. 3, if all the group members' individual answers were correct, the consensus of the same answer was 97.59%. If the group members' individual answers were incorrect, and the incorrect answers were the same, the consensus of the first discussion was 100%. There is no statistical difference between type A-S choosing the group correct answer and type D-S choosing the group incorrect answer ($\chi^2 = 0.443$, $df = 1$, $p = 0.505$). In other words, if the individual answers were the same in the group, this represented consensus on the same answer.

When individual answers are partially different, how is the group performance?—Cases with partially different individual answers were divided into three categories: more than half of the group provided the correct answer or equal number of correct and incorrect answers in the individual round (type B-P), the correct answers were provided by less than half in the individual round (type C-P), and no one in the group provided the correct answer in the individual round (type D-P). As shown by the yellow bars in Fig. 3, the group performance is higher in the order of the overall number of correct individual answers.

When all individual answers differ, will one answer lead the group answer, or will a new answer be constructed?—When all the individual-round answers within the group were different, the group decision was different depending on whether the correct response was included. When all the individual answers within a group differed, they were divided into two categories: (1) one correct answer and many different incorrect answers (type C-D), and (2) all different incorrect answers (type D-D). As shown by the green bars in Fig. 3, when the group contained a single correct response, 80.71% of them chose the correct group answer. It was higher than the expected rate based on equal weight of individual answers of 20%–25%, when 4–5 group members chose different answers.

Meanwhile, when the correct response was not included in the group before discussion, the group answer was expected to be incorrect no matter which respondent the group followed. However, in first stage of the group round, 38.78% of the respondents constructed their knowledge and could select the correct group answer without selecting one of the group member's individual-round incorrect answers. This means that when the variety of the answers increased, group performance is better for the groups, including a correct answer (type C), than the groups without a correct answer in the individual round (type D) overall. When, the performances of the groups type C-N and type D-N are compared, performance is 80.71% for type C-N, it is 38.78% for type D-N. That means, although the answers in the groups are not in a majority (all incorrect answers are different), group performance is better for the groups including a correct answer (type C) than the groups without a correct answer in the individual round (type D) overall.

V. CONCLUSION, DISCUSSION, AND IMPLICATIONS

Previous studies on peer collaboration in physics learning found that students achieved statistically better scores when they collaborate [5,15,23,26]. Additionally, the research on collaborative exams in other disciplines indicated that students achieved better and statistically significant performance scores when they collaborate [25,30,32–34]. In this study, we hypothesized that students' individual-round performances, both with correct and incorrect answers, were critical for their performance when they collaborated. According to the data analysis, based on 2047 instances of two-round collaborative testing over three semesters, the conclusions of this research are as follows.

A. Conclusion 1

The number of individual-round correct answers is important for group performance both on its own and together with the consideration of the majority and the variety of the answers.

The performance differed, depending on the number of individual responses of group members. That means, the percentage of finding the correct answer in the group round was higher based on the number of correct respondents in the preceding individual round (performance for type A > type B > type C > type D). As the number of correct answers given in the individual round decreases, the possibility of members discovering the right answer in the first stage of the group round decreases.

Moreover, the number of correct answers was considered in conjunction with the majority of correct or incorrect answers. Group performance was higher for groups with equal or more correct answers (type A and B) than for groups with less than half correct responses (type C) overall when the number of correct answers are dominant to the same incorrect answers or each different incorrect answer (performance for type A-C ≈ type B-C > type C-C); group performance is better for the groups including a correct answer than the groups without a correct answer in the individual round overall when the majority is the same incorrect answer (performance for type C-I > type D-I), and none of the answers constitute a majority in the groups (all incorrect answers are different) (performance for type C-N > type D-N).

For the variety and number of correct answers considered together, the group performance is higher for the partially same answers in the order of the number of correct individual answers in overall (performance for type B-P > type C-P > type D-P). Additionally, when the variety of answers increased, i.e., when the groups had all different answers, the performance of groups with a correct answer was higher than the groups without a correct answer in the individual round (performance for type C-D > type D-D).

The groups starting their discussion with a greater number of members who correctly answered may have more knowledge to exchange, or they may better activate relevant concepts [5,14,15,21]. The students who gave correct answers in the individual rounds may create cognitive gain and offer intelligible, plausible, and fruitful concepts that provide the conflict necessary to drive conceptual change in the group [35–37]. In addition, the students who correctly answered in the individual round may provide a zone of proximal development [2], within which their peers, who answered incorrectly in the individual round, can perform better during peer discussion, and thus find the correct answer.

The groups with few correct answers in the individual round can also gain scientific understanding of the concepts through peer discussions. The students in these groups might have misconceptions or partial knowledge of a scientific concept. These students might often begin discussions related to physics content and ask many questions during peer collaboration [18]. New questions posed by peers may cause dissatisfaction with a student's own knowledge, known as the conceptual conflict [35]. So learning by collaboration in these groups might be considered a radical change or reconstruction of knowledge rather than knowledge activation or recall in contrast to the groups with more correct answers in the individual round.

B. Conclusion 2

A majority of individual-round correct or incorrect answers contributes to group performance.

The group performance showed the highest majority with correct answers, followed by the nonmajority, a tie between correct and incorrect answers, and a majority with incorrect answers. In addition, chi-square analyses for the majority indicated that even though the number of correct answers in the group is the same, the group performance was statistically different depending on a majority in answers (performance for type B-C > type B-T; type C-C > type C-N > type C-T > type C-I; type D-N > type D-I).

When individual answers within a group are predominantly correct answers, group performance is higher in the same group type or different group types. When the nonmajority and majority in incorrect answers were observed for some groups at the same time (types C and D), group performance was better for the nonmajority—all answers were different from each other—than the incorrect majority. This performance also increases if there is one correct answer among all different answers (performance for type C-N > type D-N).

C. Conclusion 3

The variety of answers in the individual round is important for group performance.

When variety in answers was observed for some groups at the same time, irrespective of groups containing a correct answer or not, when the variety of answers increased group performance was higher. Additionally, chi-square analyses for variety indicated that even when the number of correct answers in the group was the same, the group performance was statistically different, depending on the variety (performance for type C-D > type C-P; type D-D > type D-P > type D-S). If there is one correct answer among all various answers, this performance also increased (performance for type C-D > type D-D).

The variety of answers affected the performance of co-construction of physics knowledge as a group. That means the groups with a variety of incorrect respondents—such as all different, partially different, or all the same—differed in terms of their performance in the group round. When the variety of incorrect answers in individual rounds was large, the group performance was high. The role of the diversity of group members in terms of ability, the number of students in the groups and the number of male students in the group did not significantly impact students' learning previously [27]. The results of our study indicated that the cognitive variety of group members might have a role in learning and influence discovery of the correct answer after discussion. This was also observed in the groups that did not have the correct answer in individual rounds. Consensus on the same wrong answer might be a result of the same misconception and 0% performance in the first stage of the group round might be evidence of how that unscientific knowledge was robust to change with the knowledge with limited variety. In contrast, a variety of answers might help students' organization of their knowledge by first conceptual conflict [35] and then showing the implausible and unfruitful nature of other knowledge [35–37] by their peers. This way students in these groups might present higher group performances by radical restructuring [36] with marked revision of existing knowledge by more inquiry than the others.

Through collaboration, knowledge construction in groups where there was no correct answer in the individual round was also observed in the previous studies [4–6]. The co-construction of knowledge [5] with peers provided the greatest cognitive benefits through peer discussion by increase in common knowledge for all collaborating partners [9]. Singh [5] showed this type of interaction in a calculus-based physics course, identifying how peers with incorrect answers could find the correct answers. She linked groups that found the correct answer, even when group members had given an incorrect answer in the individual round, to the articulation of ideas during the peer discussion in the group round. In addition to the social and discursive dimension of co-construction, due to the disciplinary aspect students could focus on producing an answer to the question by using the information they have when co-constructing their physics knowledge [19].

D. Conclusion 4

If the individual-round answers are not various, either all correct or all the same incorrect, the groups reach a consensus on the response.

The percentage of consensus for all correct and all the same incorrect answers was not statistically different (type A-S $\approx$ type D-S). However, the performance of the groups with all correct answers was the highest among all groups, while the performance of the groups with all incorrect answers was the lowest.

In group members with the same individual-round answers, most obtained consensus in the first-round discussion, and this was true for 100% of groups with the same incorrect answers, and 97.59% of groups with same correct answers. How students who hold metalevel discussions, presented disagreements and agreements to their peers were also identified previously [18]. 2.41% of type A members might also indicate that kind of peers in group discussions.

In two-round physics problem solving, students benefit from collaborative discussion, not just majority opinion. So including many misleading, incorrect answers to increase the variety of responses within the group, rather than focusing on one answer when asking questions might be used for solving the two-round physics problem. Another alternative would be to form a group that enhances variety based on students' individual-round responses.

ACKNOWLEDGMENTS

The authors thank the Ministry of Education of the Republic of Korea, the National Research Foundation of Korea (NRF-2020S1A5B5A16083941), The Scientific and Technological Research Council of Türkiye, Professors Dr. Eric Mazur and Dr. Kelly Miller for their support and feedback.

[1] M. E. Gredler, *Learning and Instruction: Theory into Practice* (Prentice-Hall, Upper Saddle River, NJ, 1997).

[2] L. Vygotsky, *Mind in Society* (Harvard University Press, Cambridge, MA, 1978).

[3] P. N. Johnson-Laird, The history of mental models, in *Psychology of Reasoning: Theoretical and Historical Perspectives*, edited by K. Manktelow and M. C. Chung (Psychology Press, New York, NY 2004), pp. 179–212.

[4] H. Jang, N. Lasry, K. Miller, and E. Mazur, Collaborative exams: Cheating? Or learning?, Am. J. Phys. **85**, 223 (2017).

[5] C. Singh, Impact of peer interaction on conceptual test performance, Am. J. Phys. **73**, 446 (2005).

[6] 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).

[7] R. T. Johnson and D. W. Johnson, Action research: Cooperative learning in the science classroom, Sci. Child. **24**, 31 (1986), https://www.jstor.org/stable/43166314.

[8] D. W. Johnson and R. T. Johnson, *Cooperation and Competition, Theory and Research* (Interaction Book Company, Edina, MN 1989).

[9] H. Jeong and M. T. H. Chi, Knowledge convergence and collaborative learning Instr. Sci. **35**, 287 (2007).

[10] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).

[11] K. Miller, N. Lasry, B. Lukoff, J. Schell, and E. Mazur, Conceptual question response times in Peer Instruction classrooms, Phys. Rev. ST Phys. Educ. Res. **10**, 020113 (2014).
[12] K. Miller, J. Schell, A. Ho, B. Lukoff, and E. Mazur, Response switching and self-efficacy in Peer Instruction classrooms, Phys. Rev. ST Phys. Educ. Res. **11**, 010104 (2015).
[13] T. Espinosa, K. Miller, I. Araujo, and E. Mazur, Reducing the gender gap in students' physics self-efficacy in a team- and project-based introductory physics class, Phys. Rev. Phys. Educ. Res. **15**, 010132 (2019).
[14] B. Bungum, M. V. Boe, and E. K. Henriksen, Quantum talk: How small-group discussions may enhance students' understanding in quantum physics, Sci. Educ. **102**, 856 (2018).
[15] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping, Part 1: Group versus individual problem solving, Am. J. Phys. **60**, 627 (1992).
[16] R. Gautreau and L. Novemsky, Concepts first—A small group approach to physics learning Am. J. Phys. **65**, 418 (1997).
[17] E. Mazur, *Peer Instruction: A User's Manual* (Pearson & Prentice Hall, Upper Saddle River, NJ, 1997).
[18] R. Leinonen, M. A. Asikainen, and P. E. Hirvonen, Peer discussions in lecture-based tutorials in introductory physics, Phys. Rev. Phys. Educ. Res. **13**, 010114 (2017).
[19] A. Pawlak, P. W. Irving, and M. D. Caballero, Development of the Modes of Collaboration framework, Phys. Rev. Phys. Educ. Res. **14**, 010101 (2018).
[20] M. Sabella and E. F. Redish, Knowledge organization and activation in physics problem solving, Am. J. Phys. **75**, 1017 (2007).
[21] A. K. Wood, R. K. Galloway, J. Hardy, and C. M. Sinclair, Analyzing learning during Peer Instruction dialogues: A resource activation framework, Phys. Rev. ST Phys. Educ. Res. **10**, 020107 (2014).
[22] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. V. Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69**, S12 (2001).
[23] C. S. Kalman, S. Morris, C. Cottin, and R. Gordon, Promoting conceptual change using collaborative groups in quantitative gateway courses, Am. J. Phys. **67**, S45 (1999).
[24] M. J. Giuliodori, H. L. Lujan, and S. E. DiCarlo, Peer instruction enhanced student performance on qualitative problem-solving questions, Adv. Physiology Educ. **30**, 168 (2006).
[25] M. J. Giuliodori, H. L. Lujan, and S. E. DiCarlo, Collaborative group testing benefits high- and low-performing students, Adv. Physiology Educ. **32**, 274 (2008).
[26] A. Mason and C. Singh, Helping students learn effective problem solving strategies by reflecting with peers, Am. J. Phys. **78**, 748 (2010).
[27] J. J. B. Harlow, D. M. Harrison, and A. Meyertholen, Effective student teams for collaborative learning in an introductory university physics course, Phys. Rev. Phys. Educ. Res. **12**, 010138 (2016).
[28] C. E. Wieman, G. W. Rieger, and C. E. Heiner, Physics exams that promote collaborative learning, Phys. Teach. **52**, 51 (2014).
[29] D. Bloom, Collaborative test taking: Benefits for learning and retention, Coll. Teach. **57**, 216 (2009).
[30] R. Cortright, H. L. Collins, D. Rodenbaugh, and S. DiCarlo, Student retention of course content is improved by collaborative-group testing, Adv. Physiology Educ. **27**, 102 (2003).
[31] S. Kapitanoff, Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance, Active Learning Higher Educ. **10**, 56 (2009).
[32] B. H. Gilley and B. Clarkston, Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students, Res. Teach. Engl **43**, 83 (2014).
[33] J. F. Zipp, Learning by exams: The impact of two-stage cooperative tests, Teaching Sociology **35**, 62 (2007).
[34] H. Leight, C. Saunders, R. Calkins, and M. Withers, Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class, CBE Life Sci. Educ. **11**, 392 (2012).
[35] P. W. Hewson, A conceptual change approach to learning science, Eur. J. Sci. Educ. **3**, 383 (1981).
[36] S. Carey, *Conceptual Change in Childhood* (MIT Press, Cambridge, MA 1985).
[37] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).