Educ Inf Technol (2018) 23:2323–2340
https://doi.org/10.1007/s10639-018-9717-3

Investigating the effects of peer instruction on preservice mathematics teachers' achievements in statistics and probability

Yusuf Ziya Olpak[1] **· Serdal Baltaci**[2] **·**
Muhammet Arican[2]

Received: 21 February 2018 / Accepted: 17 April 2018 / Published online: 4 May 2018

Abstract This study investigated the effects of two different accountability scoring mechanisms (ASMs), which were used during the peer instruction (PI) process, on preservice middle school mathematics teachers' (PSTs) achievements in statistics and probability. In the spring semester of 2016–2017 academic year, 46 third-year PSTs, who had been attending a statistics and probability course, participated in the study. Based on their pre-test scores, the PSTs were randomly divided into two equally achieving groups (Group 1 and Group 2). The data of this study were collected using an academic achievement test and PI and course evaluation forms. A learning management system (LMS), which was a web-based application designed by the first author, was used in collecting the PSTs' responses to the given conceptual questions. Two different ASMs were used in calculating the PSTs' PI scores. The findings showed that the PSTs in Group 1 obtained significantly higher overall academic, PI, and post-test scores than the PSTs in Group 2. Therefore, the PSTs' PI and overall academic achievement scores differed based on the ASM used. Thus, using an ASM during a PI process found to be effective in increasing the PSTs' engagement in the peer discussion. Finally, the analysis of the PSTs' opinions regarding to the PI, learning process, and learning environment indicated their overall satisfaction.

Keywords Accountability scoring mechanism · Mathematics achievement · Peer instruction · Preservice teacher education · Statistics and probability

✉ Muhammet Arican
muhammetarican@gmail.com

[1] Computer Education and Instructional Technology Department, Ahi Evran University, Kirsehir, Turkey

[2] Mathematics and Science Education Department, Ahi Evran University, Kirsehir, Turkey

1 Introduction

The development of technology provokes changes and transformations in educational environments as well as in every area of daily-life. Hence, it is important for students to have twenty-first century skills during the lifelong learning process. The International Society for Technology in Education (ISTE 2017) determined these twenty-first century skills as follows: Empowered learner, digital citizen, knowledge constructor, innovative designer, computational thinker, creative communicator, and global collaborator. The higher education, which is the last stage of formal education, is very important in acquiring these skills since at the end of this stage, individuals are qualified to be owners of professions. Therefore, it is important for preservice teachers (PSTs) to have these skills mentioned above because acquiring them will eventually make an impact on their future students and on developing qualified generations. During the acquisition process of these skills, rather than a theory-oriented teacher-centered approach, an application-oriented student-centered approach should be preferred. Furthermore, methods and techniques should allow PSTs to be active in their learning experiences.

One of the methods that can be used in developing PSTs' twenty-first century skills is the PI method developed by Eric Mazur. According to Mazur (1997), it is very difficult for teachers to draw students' attention during an entire lesson, especially in large classes. Students often view large class lectures as impersonal (MacArthur and Jones 2008), and these lectures are usually performed as a monologue in front of passive students (Biggs 1996; Mazur 1997). Besides, studies have revealed that students benefit from discussing their ideas with one another in class (Mazur 1997; Michinov et al. 2015). Although the PI method was developed in a physics course, it has been applied in various areas such as mathematics (e.g., Latulippe 2016; Miller 2013), computer science (e.g., Chou and Lin 2015; Lee et al. 2013), chemistry (e.g., Michinov et al. 2015; Morice et al. 2015), and engineering (e.g., Adawi et al. 2016; Arteaga and Vinken 2013).

Studies (e.g., Fagen et al. 2002; Kay and LeSage 2009; Mazur 1997) showed that the PI method has some limitations in terms of instructors and students. For instance, Kay and LeSage (2009) pointed out two of the difficulties encountered by the instructors as generating effective conceptual questions and establishing a learning system to be used during the question-answer process. Similarly, they stated the difficulties encountered by the students as negative reactions to being monitored, not understanding course contents when discussed from multiple perspectives, increased confusion, and adaptation to a new learning method.

Regarding the advantages, Mazur (1997) stated that the PI method ends the monotony of a passive narrative, and students receive instant feedbacks on their responses, which also help instructors to detect students' understanding of a subject. Cortright et al. (2005) described the PI as a cooperative learning technique that promotes critical thinking, problem solving, and decision-making and enhances students' meaningful understanding. Moreover, the PI improves students' performance in qualitative problems (Giuliodori et al. 2006), and teachers can create active and effective learning environments in crowded classrooms with little effort (Rao and DiCarlo 2000).

Although the PI is a flexible student-centered method, which may vary for different learning scenarios, it mainly consists of the lecture and question-answer processes

(Chou and Lin 2015; Crouch et al. 2007; Mazur 1997). The multiple choice questions were used in the question-answer process of the PI are referred to as ConcepTests. Mazur (1997) summarized the general steps of a ConcepTest as follows: A question is posed by the instructor, students are given time for thinking, students record their individual answers (optional), they try to convince their group-mates through peer discussion, they record their revised answers (optional), the instructor provide feedbacks by tallying students' responses and explaining the correct answer. However, as reported by Dancy and Henderson (2010), less than 12.8% of the instructors use the PI method as it was originally designed to be implemented.

The questions in a ConcepTest can be presented to the students via a blackboard, a projector or in oral form (Chou and Lin 2015). Depending on the purpose of the study and features of the learning environment, students can provide answers showing their hands (Chou and Lin 2015; Mazur 1997), using scanned forms (Mazur 1997), handheld computers (Mazur 1997), electronic voting systems (Draper and Brown 2004), and classroom response systems (Bruff 2009). Although Mazur (1997) considered the PI's success to be independent of the feedback method and technological resources, the use of classroom response systems with technological possibilities reported to have some advantages on improving students' learning experiences (Blasco-Arcas et al. 2013; Yourstone et al. 2008).

Kay and LeSage (2009) reported that allowing students to use remote control devices when answering multiple choice questions provided improvements in the following conditions: Classroom environment (increase in student attendance, levels of attention, classroom participation, and engagement), learning (increase in interaction, discussion, contingent teaching, quality of learning, and learning performance), and assessment (increase in the number of feedbacks either formative or normative). Similarly, Lucas (2009) used the PI method in a calculus class to encourage student cooperation and to promote their activate participation in the learning processes. The results of the study showed that the use of i-clickers in the PI sessions not only made learning an effective and fun process, but also enhanced students' learning since they considered their peers as valuable resources. Furthermore, Miller (2013) collected preservice elementary teachers' opinions on the effectiveness of using clickers during a mathematics course. The PSTs stated that the inclusion of clickers assisted them learning and understanding the course materials.

The PI literature showed researchers' different approaches in creating effective collaborative learning environments (Chou and Lin 2015; Len 2007; Trout et al. 2014). For example, Trout et al. (2014) conducted all the classroom lectures online, assigned readings to the students, attendance was not mandatory, and the conceptual questions and students' answers were posted online after each class session for review. Moreover, Len (2007) investigated the effects of different award systems, which were provided during the PI session to increase student motivation, on students' learning and behaviors. Furthermore, Chou and Lin (2015) used discussion partner assignment and ASMs to promote students' discussion. Their findings showed that some students still did not participate in the classroom discussions. Hence, they reported the students' hesitation in participating discussions to be related with their lack of motivation, being afraid of participating in discussions, and disconcerting to discuss with unfamiliar group-mates.

When mathematics education literature was examined, we recognized that there is a lack of knowledge on the effects of the PI on PSTs' understanding of statistics and probability. Statistics and probability are regarded as two of the challenging topics for students and PSTs. Bulut et al. (1999) reported that the large majority of students have a negative attitude towards statistics and probability, they have difficulty understanding problems and memorize formulas to solve these problems, and teachers have difficulty using appropriate teaching methods. Moreover, Celik and Gunes (2007) observed students' difficulties in explaining the reasons behind their correct answers for the statistics and probability questions. Similarly, the second author of this study, who has been teaching this course for several years, observed such difficulties of the PSTs'. In our view, these difficulties have been arising because PSTs usually not being able to associate what they have learned in statistics and probability with their prior knowledge. Furthermore, they are not given enough opportunities for collaborative learning experiences and for applying what they have learned in daily-life and in other courses to the statistics and probability. Thus, we decided to conduct this study in an undergraduate statistics and probability course to overcome difficulties discussed above.

Considering the advantages of technological resources on PSTs' learning, the LMS, a web-based application designed by the first author, was used during the question-answer process of the PI. The PSTs used the question-answer module of the LMS to answer active conceptual questions and received information (i.e., question, choices, correct answer, their first and second responses and group-mate's first and second responses, correctness of these responses, and points that they and group-mate received from these responses) on already answered questions. The PSTs were divided in two equally achieving groups, and in each group, they were allowed to choose their group-mates, which we expected to increase their motivation and participation in the class discussions. In two groups, we used the ASM in scoring the PSTs' responses to the given conceptual questions. In addition, we collected the PSTs' opinions regarding to the effectiveness of the PI method. Therefore, we investigated the following research questions:

1. How do preservice mathematics teachers' academic achievement in statistics and probability differ based on the ASM used?
2. What are the preservice mathematics teachers' opinions in two groups regarding to the PI, learning process, and learning environment?

2 Methodology

This study was developed following an explanatory research design model. In the explanatory research design model, researchers use quantitative methods to analyze data and refine their findings using qualitative methods (Fraenkel and Wallen 2006). Following the explanatory research design model, in this study, both quantitative and qualitative methods used in analyzing data. In the following pages, participants, data collection tools, the learning management system (LMS), and data collection and analysis procedures were explained.

2.1 Participants

During spring semester of 2016–2017 academic year, 52 PSTs, who enrolled at the Mathematics and Science Education Department of a public university in Turkey, participated in the study. A purposive sampling technique was used in the selection of the PSTs. All the PSTs were in their third years, who were specialized at middle school mathematics education, and were attending to an Introductory Statistics and Probability course. The course content included distributions of probability (i.e., Normal, Binomial, Poisson, and Hypergeometric), sampling theory, and estimation theory. All of the PSTs stated that they were not familiar with the PI method. In addition, they did not have any online learning experience and did not use the LMS in their classes until this study.

To determine the PSTs' prior knowledge on the course topics and to assign them into two groups, we administered an academic achievement test developed by Arican and Kuzu (Diagnosing preservice teachers' understanding of statistics and probability: Developing a test for cognitive assessment, unpublished manuscript) during the first week of the class. Because the PSTs' pre-test scores were low and similar, we randomly assigned them into two groups (Group 1 and Group 2). Both groups initially included 26 PSTs in each. Because six PSTs did not regularly attend to the classes, the data analysis were carried out using responses collected from 46 PSTs (41 Female and 5 Male). Group demographics are reported in Table 1. There were 22 female (91.67%) and two male (8.33%) PSTs in Group 1. Whereas, Group 2 consisted of 19 female (86.36%) and three male (13.64%) PSTs. Because of the university's demographics and female students' tendencies in attending teacher education programs, the study included more of the female PSTs than the male PSTs. The PSTs were not permitted to switch between Group 1 and Group 2 and were instructed not to work with members of the other groups while in the class. However, within each group, the PSTs were permitted to select their group-mates. Hence, within Group 1 and Group 2, two PSTs formed the sub-groups and worked together until the last week of the semester. In the literature, the consequences of group sizes on the effectiveness of the PI method on students' achievements are not well explored. Nevertheless, Morice et al. (2015) noted that permitting students to work in group sizes of two to four precluded social loafing and improved their learning. Therefore, in this study, two PSTs in both groups allowed to work together.

During the question-answer process, two different ASMs, which were presented in Table 2, were used in calculating the PSTs' PI scores. The purpose of using the ASM was to increase the PSTs' course attendance and to encourage them for peer discussions. In Group 1, the PSTs received 40% of the total score in their first correct responses. If they still chose the correct option after discussing their answers with group-mates, they received another 30% of the score. Furthermore, they received an extra 30% of the score if their group-mates also chose the correct option in their second responses. Whereas, in Group 2, the PSTs' scores were calculated taking 50% of their first correct responses, 40% of their second correct responses, and they received an extra 10% of the score based on their group-mates' correct responses. The PSTs' PI scores were automatically calculated by the LMS.

Table 1 Group demographics

Group	Gender	N	%
Group 1	Female	22	91.67
	Male	2	8.33
Group 2	Female	19	86.36
	Male	3	13.64

Table 2 The accountability scoring mechanisms for group 1 and group 2

	Group 1		Group 2	
	First Answer	Second Answer	First Answer	Second Answer
Student	40%	30%	50%	40%
Discussion Partner	0%	30%	0%	10%

2.2 Data collection tools

As stated earlier, an academic achievement test, which was designed by Arican and Kuzu (unpublished manuscript), was used as a pre-test in determining the PSTs' prior knowledge in statistics and probability. The test included 15 multiple-choice and five open-ended questions. The same test was given to the PSTs as a post-test at the end of the semester to calculate the differences between mathematics achievements of the two groups. In addition to the mathematics achievement test, the PSTs' PI session scores, which they obtained answering conceptual questions during the question-answer process, were also used in determining their overall academic achievements. Table 3 reported the contribution of the PI and post-test scores on the PSTs' overall course achievements.

The conceptual questions were prepared by the course instructor, the second author of this study. These questions were controlled by two mathematics teachers and two mathematics educators for their contents and relevance to the PSTs' knowledge levels. Furthermore, the PSTs were given two evaluation forms to understand their overall satisfactions from the PI method, learning process, and learning environment. The PSTs' evaluations on the PI method were collected using the "Peer Instruction Student Evaluation Form" designed by Olpak et al. (2017). Finally, their evaluations on the learning process and environment were collected by the "Course Evaluation Form," which was designed by the authors.

2.3 The learning management system

The PSTs used the LMS to access the conceptual questions through their web browsers. In the LMS, there were two different permission groups (learner and instructor). All users accessed the system entering their usernames and passwords. After entering the correct usernames and passwords, the application directed the instructor and learners to a "Home" screen (Fig. 1). There were following menus in the LMS home screen: The "Lessons" menu enabled the instructor and learners to access the conceptual questions and syllabus. While the "Messages" menu allowed the learners to communicate asynchronously between group-mates and instructor, the "Blog" menu allowed whole class asynchronous communications including the instructor. On the other hand, the "Chat"

Table 3 The contribution of the PI and post-test scores on the overall academic achievement

Assessment	Contribution (%)
PI scores	30
Post-test scores	70

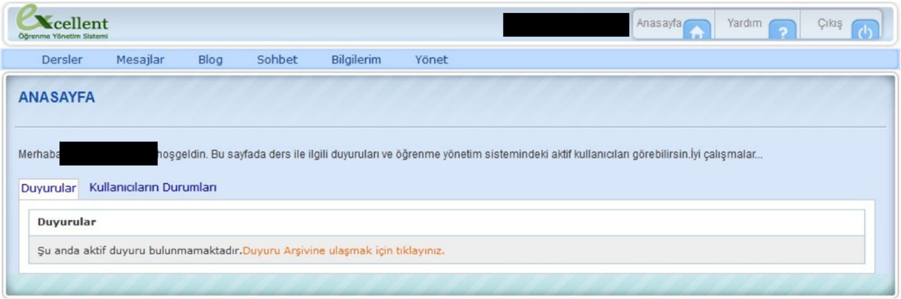

Fig. 1 A screen shot from the "Home" screen of the LMS

menu was designed for synchronous communication between group-mates and instructor. Furthermore, using the "My Info" menu, the instructor and learners accessed information about themselves and changed their passwords if needed.

In addition to these five menus, the LMS designed for the instructor included the "Manage" menu. Using the "Manage" menu, the instructor was able to make announcements, share files in different formats, arrange study groups among the learners, prepare multiple-choice questions, edit questions that were already in the system, activate questions during the question-answer process, monitor the PSTs' participation in course activities, and check their average PI scores. In addition, the PSTs used the "Question-Answer" module (Fig. 2), which was located under the "Lessons" menu, to response active questions and receive information (i.e., questions, choices, correct choice, first and second responses and group-mates' first and second responses, correctness of these responses, and scores that they and their group-mates received from these responses) on already answered questions.

2.4 The data collection and analysis

The study was conducted in a statistics and probability course. The course was 4-h long each week and divided into two 2-h sessions (theory and application). During the first

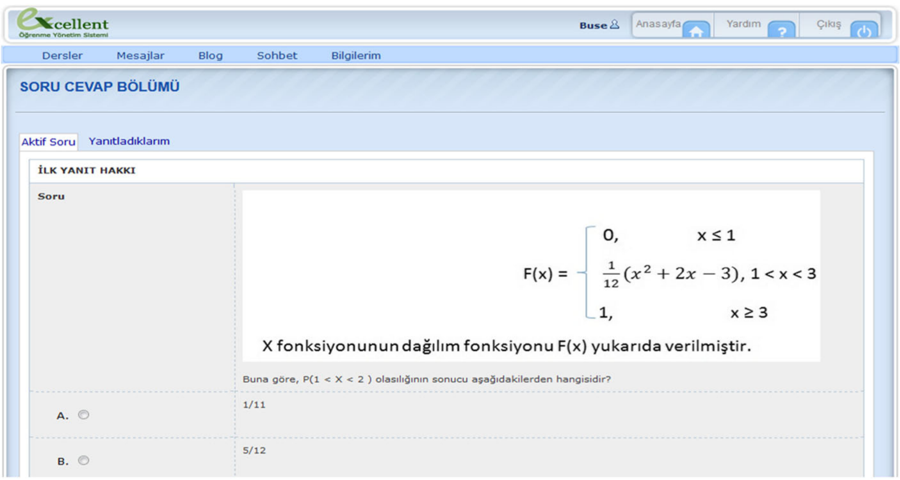

Fig. 2 A screen shot from the "Question-Answer" module

week, the instructor provided information about the course content and shared his expectations. Next, in the application session, the pre-test was conducted that measured the PSTs' prior knowledge on the course topics. Beginning from the second week, the PSTs in two groups attended together to the theory session in a regular classroom. Later, both groups attended to two separate application sessions in a computer laboratory. The second week, in the application session, the instructor briefly explained what a PI method was and introduced the question-answer process to the PSTs in both groups. In each week, after the instruction on that week's topics in the theory session, the PSTs answered conceptual questions, which were activated by the instructor, using the question-answer module in the LMS.

The PSTs' responses to the pre- and post-tests and the data collected during the PI question-answer process were analyzed using the Statistical Package for Social Sciences (SPSS) in which all hypotheses were tested at a .95 confidence level ($p = .05$). The PSTs' overall academic achievements were calculated using their PI and post-test scores (see Table 3). An Independent Samples T-Test was used to understand whether the PSTs in Group 1 and Group 2 were significantly differed based on their PI scores and overall academic achievements. Furthermore, a One-way Analysis of Covariance (ANCOVA) test was used to examine whether the PSTs' post-test scores, which were adjusted according to their pre-test scores, significantly differed in both groups or not. In addition, the PSTs' evaluations on the "Peer Instruction Evaluation Form" were reported using the descriptive statistics, and an Independent Samples T-Test was conducted to understand the differences between the PSTs' evaluations in both groups. Finally, the PSTs' evaluations on the "Course Evaluation Form" were analyzed using the qualitative content analysis method, and the findings were reported using the descriptive statistics.

3 Findings

The Independent-Samples T-Test was used to investigate if the PSTs' PI and overall academic achievement scores differed significantly in two groups. The T-Test results were presented in Table 4. According to Table 4, in Group 1, the PSTs' average PI and overall academic achievement scores were calculated as 78.57 (5.87) and 77.82 (9.20), respectively. On the other hand, in Group 2, the PSTs' average PI and overall academic achievement scores were calculated as 67.83 (12.85) and 69.82 (9.00), respectively. The T-Test results showed that the PSTs' PI [$t(44) = 3.695$, $p < .05$] and overall academic achievement scores [$t(44) = 2.973$, $p < .05$] in two groups differed

Table 4 The independent-samples t-test results for the PSTs' PI and academic scores

Scores	Group	N	$\overline{X}$	Std	df	t	p
PI Scores	Group 1	24	78.57	5.87	44	3.695	.001
	Group 2	22	67.83	12.85			
Overall Academic	Group 1	24	77.82	9.20	44	2.973	.005
Achievement Scores	Group 2	22	69.82	9.00			

significantly according to the ASM used. Hence, the analysis suggested that the PSTs in Group 1 obtained higher PI and overall academic achievement scores than the PSTs in Group 2. The η^2 values of the PSTs' PI and overall academic achievement scores were calculated as .24 and .17, respectively. These values suggested that approximately 24% of the variance observed in the PI scores and 17% of the variance observed in the overall academic achievement scores were explained by the ASM used.

In addition to comparing the PSTs' PI and overall academic achievement scores, we also compared their post-test scores. The PSTs' post-test scores were adjusted according to their pre-test scores. We presented the post-test and adjusted post-test mean scores in Table 5. According to Table 5, in Group 1, the PSTs' mean post-test and adjusted mean post-test scores were calculated as 77.5 and 77.71, respectively. On the other hand, in Group 2, the mean and adjusted mean scores were calculated as 70.68 and 70.45, respectively.

To understand if there was a meaningful difference in two groups' post-test scores, an ANCOVA analysis was conducted (Table 6). Table 6 showed a meaningful difference in the post-test scores of two groups [$F_{(1, 43)} = 5.379$, $p = .025$] suggesting that the PSTs in Group1 performed better in the post-test than the PSTs in Group 2.

3.1 The preservice teachers' responses to the evaluation forms

The last week of the course, we collected the PSTs' responses to the "Peer Instruction Student Evaluation" and "Course Evaluation" forms. The "Peer Instruction Student Evaluation Form" consisted of 25 likert-type items in three sub-dimensions: Student evaluations regarding to the PI method, student evaluations regarding to the conceptual questions, and student evaluations regarding to the peer discussions. The factor loadings of the items ranged between .90 and .93, and the Cronbach's alpha value was measured as .92. Hence, these measures suggested that the "Peer Instruction Student Evaluation" form was a reliable data collection tool. We presented the PI student evaluation form and mean scores of the PSTs' responses in Table 7.

The frequencies, means, and standard deviations related to the PSTs' evaluations were presented in Table 7. Table 7 showed that in both groups, except a few items, the PSTs' evaluations regarding to the PI method were mostly positive. In addition to providing the descriptive statistics, we conducted an ANOVA analysis for independent samples to understand whether the PSTs' evaluations differed significantly in both groups. The ANOVA analysis indicated that the PSTs' evaluations were significantly differing for Item 4 (Peer instruction method was enjoyable), [$F_{(1,44)} = 4.231$, $p = .046 < .05$], Item 12 (When I consider all the activities in the course, I think the allocated time for the peer instruction method was sufficient) [$F_{(1,44)} = 4.226$, $p = .046 < .05$], Item 14 (I think peer instruction method was useful) [$F_{(1,44)} = 4.716$, $p = .035 < .05$], and Item 21 (The discussion level of the peer instruction was high) [$F_{(1,44)} = 12.614$, $p = .001 < .05$]. In Item 4, although the mean scores of the PSTs' evaluations in both groups were positive, more

Table 5 The descriptive statistics of the post-test scores

Groups	N	Mean	Adjusted Mean
Group 1	24	77.50	77.71
Group 2	22	70.68	70.45

Table 6 The ANCOVA results for adjusted post-test mean scores

Source	Type III Sum of Squares	df	Mean Square	F	p
Pre-test (Reg.)	504.627	1	504.627	4.511	.039
Group	601.749	1	601.749	5.379	.025
Error	4810.144	43	111.863		
Corrected Total	5848.369	45			

PSTs in Group 1 tended to offer positive evaluations. We found a similar result for Item 12. On the other hand, in Items 14 and 21, the PSTs in Group 2 provided negative evaluations in which the mean scores were 2.27 and 2.50, respectively. This result suggested that the PSTs in Group 2 found the PI method as less useful and the level of peer discussions as insufficient.

The "Course Evaluation Form" consisted of five items. In Item 1, the PSTs evaluated their overall satisfaction from the statistics and probability course on a scale that included points between 1 and 5, in which 5 points indicated a very strong satisfaction and 1 point indicated a very weak satisfaction. The PSTs also provided explanations for their evaluations. When discussing the PSTs' explanations, we reported the PSTs in Group 1 as PST1-PST24 and reported the PSTs in Group 2 as PST25-PST46. The frequency of points and percentages were presented in Table 8. Table 8 showed that all PSTs in Group 1 gave a score of 3 points and above with an average score of 4. Similarly, except one PST, who gave a score of 1 point, the remaining PSTs gave a score of 3 points and above with an average score of 3.9. Therefore, the PSTs' overall satisfaction from the course was very much similar.

Regarding the PSTs' explanations, PST44, who gave 5 points, emphasized that because of the PI method, he was able to repeat topics that they learned in previous lesson and paid attention to the group work. Moreover, he noted that the knowledge they learned in class became more permanent. Furthermore, PST2, who gave 4 points, stated that she did not have to spend too much time on studying exams. However, she added that more time should have given for answering conceptual questions. On the other hand, PST35, who gave 1 point, complained about her group-mate not regularly attending to the course, which decreased her overall PI score, and so she stated her dissatisfaction about this situation:

> *PST35:* My group-mate did not regularly attend to the course, and this situation demoralized me. Hence, I did not have motivation for preparing course topics before coming to class. In addition, the course attendance was required, and this was disturbing me.

In the second question, the PSTs evaluated their overall satisfaction from the course activities. As indicated by Table 9, all PSTs in Group 1 and Group 2 gave a score of 3 points and above with an average score of 4 and 4.18, respectively. Most of the PSTs in two groups gave a score of 4 points, and this result showed their satisfaction from the course activities. The PSTs in two groups usually

Table 7 The PI student evaluation form

Student Evaluations Regarding to the PI Method	Group	Strongly Disagree ←---→ Strongly Agree (f)					Mean	Std. Deviation	Std. Error
		1	2	3	4	5			
1. Peer instruction method was clear.	G1	–	–	1	7	16	4.63	.576	.118
	G2	–	1	1	11	9	4.27	.767	.164
2. Peer instruction method was easy to follow.	G1	–	–	2	9	13	4.46	.658	.134
	G2	–	–	6	9	7	4.05	.785	.167
3. Peer instruction method was interesting.	G1	–	1	6	5	12	4.17	.963	.197
	G2	1	1	4	9	7	3.91	1.065	.227
4. Peer instruction method was enjoyable.	G1	–	–	6	7	11	4.21	.833	.170
	G2	–	3	8	5	6	3.64	1.049	.224
5. Peer instruction method helped me better understand the course topics.	G1	–	–	4	12	8	4.17	.702	.143
	G2	1	–	4	10	7	4.00	.976	.208
6. Peer instruction method helped me move beyond my previous level of knowledge.	G1	–	–	8	13	3	3.79	.658	.134
	G2	–	–	7	11	4	3.86	.710	.151
7. Peer instruction method helped me assess my level of knowledge regarding to the course topics.	G1	–	–	2	11	11	4.38	.647	.132
	G2	–	–	1	14	7	4.27	.550	.117
8. Immediate feedback with the peer instruction method helped me complete my deficiencies.	G1	1	3	7	8	5	3.54	1.103	.225
	G2	1	1	9	6	5	3.59	1.054	.225
9. Peer instruction method has increased my confidence in doing course topics.	G1	–	–	7	13	4	3.88	.680	.139
	G2	1	1	7	12	1	3.50	.859	.183
10. Peer instruction method increased my participation in class.	G1	–	1	8	6	9	3.96	.955	.195
	G2	1	1	4	11	5	3.82	1.006	.215
11. Peer instruction method increased my motivation towards the course.	G1	–	1	2	15	6	4.08	.717	.146
	G2	–	–	5	13	4	3.95	.653	.139
12. When I consider all the activities in the course, I think the allocated time for the peer instruction method was sufficient.	G1	–	–	8	11	5	3.88	.741	.151
	G2	1	–	11	9	1	3.41	.796	.170
13. I think it was difficult to apply the peer instruction method.	G1	6	11	6	1	–	2.08	.830	.169
	G2	4	14	4	–	–	2.00	.617	.132
14. I think peer instruction method was useful.	G1	2	7	7	4	4	3.04	1.233	.252
	G2	6	8	6	–	2	2.27	1.162	.248
15. I think peer instruction method should be used in other courses as well.	G1	–	–	7	13	4	3.88	.680	.139
	G2	2	–	9	9	2	3.41	1.008	.215
16. I think peer instruction method was educationally attractive.	G1	1	3	6	10	4	3.54	1.062	.217
	G2	2	1	5	12	2	3.50	1.058	.226

Table 7 (continued)

Student Evaluations Regarding to the Conceptual Questions	Group	Strongly Disagree ⟵⟶ Strongly Agree (f) 1	2	3	4	5	Mean	Std. Deviation	Std. Error
17. The questions posed in the question-answer process of the peer instruction method increased my interest.	G1	–	3	4	10	7	3.88	.992	.202
	G2	–	2	5	11	4	3.77	.869	.185
18. The questions posed in the question- answer process of the peer instruction method made it easier to understand the important points about the topics.	G1	–	1	5	10	8	4.04	.859	.175
	G2	–	2	5	9	6	3.86	.941	.201
19. The time allocated for the questions posed in the question-answer process of the peer instruction method was sufficient.	G1	1	1	4	13	5	3.83	.963	.197
	G2	1	2	3	11	5	3.77	1.066	.227
20. The level of difficulty of the questions posed in the question-answer process of the peer instruction method was appropriate for my level.	G1	–	2	8	5	9	3.88	1.035	.211
	G2	–	2	2	15	3	3.86	.774	.165
Student Evaluations Regarding to the Peer Discussions	Group	Strongly Disagree ⟵⟶ Strongly Agree (f) 1	2	3	4	5	Mean	Std. Deviation	Std. Error
21. The discussion level of the peer instruction was high.	G1	–	3	8	6	7	3.71	1.042	.213
	G2	6	5	7	2	2	2.50	1.263	.269
22. I actively participated in discussions during the peer instruction.	G1	4	13	4	3	–	2.25	.897	.183
	G2	3	8	9	2	–	2.45	.858	.183
23. I liked expressing my ideas during discussions in the peer instruction process.	G1	–	–	4	9	11	4.29	.751	.153
	G2	–	1	3	12	6	4.05	.785	.167
24. The peer instruction method enabled me being aware of the ideas of my group-mates.	G1	1	1	6	9	7	3.83	1.049	.214
	G2	–	–	12	6	4	3.64	.790	.168
25. I liked to see different perspectives during the peer instruction process.	G1	–	1	3	11	9	4.17	.816	.167
	G2	–	–	6	10	6	4.00	.756	.161

G1, Group 1; *G2*, Group 2

Table 8 The PSTs' overall course satisfaction

	Points	1	2	3	4	5
Group 1	Frequencies	0	0	3	18	3
	Percentages	0	0	12.5	75	12.5
Group 2	Frequencies	1	0	3	14	4
	Percentages	4.5	0	13.6	63.6	18.2

Table 9 The PSTs' satisfaction on the course activities

	Points	1	2	3	4	5
Group 1	Frequencies	0	0	4	16	4
	Percentages	0	0	16.7	66.6	16.7
Group 2	Frequencies	0	0	2	14	6
	Percentages	0	0	9.1	63.6	27.3

provided similar explanations. For instance, PST13, who gave 4 points, stated that the PI method increased her interaction with the group-mate. Similarly, being able to complete weekly reviews and answering questions from previous weeks facilitated her understanding of the course topics:

> *PST13:* I am very pleased with the activities. It was a successful practice for me. The PI activities increased my interaction with the group-mate, reviewing each week's questions and answering these questions again allowed us to better understand the course topics. My only concern was that laboratory and internet connection could be better. That is why I gave 4 points.

In Item 3, the PSTs evaluated their overall performance. Table 10 showed that the PSTs in Group 1 tended give higher points for their performance than the PSTs in Group 2. Moreover, all the PSTs in both groups gave themselves a score of 3 points or above. In Group 1 and Group 2, the average of evaluations were calculated as 4.66 and 3.77, respectively. This result suggested that the PSTs in Group 1 were more satisfied from their overall performance than the PSTs in Group 2.

When the PSTs' explanations were examined, they generally evaluated their overall performance as good but criticized themselves at some points. The explanations of PST6 and PST25, who gave 5 and 3 points, respectively, were provided below:

> *PST6:* I felt obliged to attend all classes, and I had to prepare every week since we solved problems using the LMS. Hence, I was able to understand topics that I was weak, and it was good to work on my own. I was able to persuade my group-mate during our weekly group discussions.
>
> *PST25:* I had to review course topics every week, but I did not do it. I could have done better. When I reviewed some questions later, I was able to find some of my mistakes. That is why my score is 3.

Table 10 The PSTs' scores for their own performance

	Points	1	2	3	4	5
Group 1	Frequencies	0	0	2	13	9
	Percentages	0	0	8.3	54.2	37.5
Group 2	Frequencies	0	0	8	11	3
	Percentages	0	0	36.4	50	13.6

Table 11 The most liked and disliked features of the course

		Features	f
Group 1	Liked	The PI was an effective method that facilitated my learning.	11
		The PI was an enjoyable method that increased my interest towards to the course topics.	9
		Preparing every week before attending to the class.	8
		The course attendance being necessary.	8
		The course discussion environment.	5
	Disliked	The drop in the PI scores in the absence of group-mates.	16
		Preparing every week before attending to the class.	12
		Computer and internet connection related problems.	10
		The course start time.	9
		The course attendance being necessary.	3
Group 2	Liked	The increase in the course attendance.	10
		The increase in motivation towards to the course content.	6
		The course discussion environment.	6
		Repeating the course topics by answering conceptual questions.	4
		Being able to notice topics that we are weak.	4
	Disliked	Waiting for group-mates to respond.	12
		The course start time.	10
		Spending too much time on solving some questions.	8
		Computer and internet connection related problems.	5

In Item 4 and Item 5, the PSTs reported their most liked and disliked features of the course, respectively. We presented the most appearing features in Table 11. Because the PSTs were allowed to state more than one feature of the course, the total number of liked and disliked features exceeded the total number of PSTs in each group. Regarding the most liked features of the course, 11 PSTs in Group 1 stated that they liked the PI and mentioned it as an effective method facilitating their learning of the course topics. Similarly, nine PSTs evaluated the PI as an enjoyable method and pointed out that it increased their interest towards to the topics. On the other hand, 16 PSTs in Group 1 stated the drop in their PI scores in the absence of group-mates as the most disliked feature of the course. Moreover, although eight PSTs stated that they liked preparing every week before attending to the class, 12 PSTs stated that they did not like this feature. In addition, 10 PSTs stated computer and internet connection related problems as one of the most disliked features of this course. In Group 2, 10 PSTs stated the increase in the course attendance, and six PSTs stated the increase in their motivation towards to the course content as the most liked features. Furthermore, six PSTs stated that they liked the course discussion environment. On the contrary, 12 PSTs stated that they did not like waiting their group-mates to respond, and 10 PSTs stated they did not like the course starting time.

In Item 6, the PSTs reported what they would like to change about the course. Table 12 showed that the PSTs in both groups suggested changing the class starting time and elimination of computer and internet connection related problems. On the other hand, while the PSTs in Group 1 also suggested changing the attendance being

Table 12 The PSTs' Suggestions for Change

	Suggestions for Change	f
Group 1	The course start time.	12
	The elimination of computer and internet connection related problems.	10
	The lesson hours should be increased.	8
	The attendance being necessary.	6
	The class-size should be reduced.	5
Group 2	The scoring mechanism (i.e., the ASM) used in calculating PI scores.	12
	The course start time.	10
	My group-mate.	6
	The elimination of computer and internet related problems.	4

necessary and class-size, the PSTs in Group 2 suggested changing the ASM and their group-mates. Some PSTs' responses on these suggestions were presented in below:

> *PST3:* I would prefer a group-free, individual work and our answers being visible immediately after we submit them. I think, the starting time of the course was very late, so I could have started it earlier. The attendance was another problem for me. There was also a need for eliminating problems related with computer and internet connection.
> *PST27:* I think, I would have changed the conditions for the absenteeism at least for some occasions because I was always worried about it, and this effected my performance and so my group-mate's PI score. Hence, I would like to change the scoring mechanism. I would design it in a way that my absence does not affect the group-mate's PI score.

4 Discussion and conclusions

In the first research question, we investigated the PSTs' achievements in statistics and probability. Hence, we compared two groups according to the differences between the PSTs' overall academic achievement, and PI and post-test scores. We calculated the PSTs' PI and overall academic scores using the ASMs presented in Table 2 and Table 3, respectively. The t-test results showed that the PSTs in Group 1 obtained higher PI and overall academic achievement scores than the PSTs in Group 2. Similarly, the ANCOVA analysis showed that the PSTs in Group 1 obtained significantly better scores in the post-test than the PSTs in Group 2. In Group 1, a PST's correct response in the second time answering a conceptual question contributed to the group-mate's total score by 30%, which was 10% in Group 2. Therefore, the PSTs in Group 1 appeared to benefit from the peer discussions more than the PSTs in Group 2. Thus, in Group 1, the ASM used in calculating the PSTs' PI scores was more effective than the ASM used in Group 2 in terms of motivating the PSTs towards peer discussions, which eventually affected their performance in course activities. Similarly, Chou and Lin (2015) reported that the ASM was increasing student engagement during classroom discussions.

In our second research question, we investigated the PSTs' opinions regarding to the PI, learning process, and learning environment. Two forms were used in collecting the PSTs' opinions: The "Peer Instruction Student Evaluation Form" and "Course Evaluation Form." The "Peer Instruction Student Evaluation Form" was used in understanding the PSTs' opinions regarding the PI method. In our analysis, we examined whether the PSTs' opinions on the PI method differed based on the study groups (i.e., Group 1 and Group 2). Except a few items, Table 7 showed the PSTs' overall satisfaction in two groups from the PI method. Generally, the PSTs in Group 1 tended to provide higher points to the PI evaluation items than the PSTs in Group 2. This finding supported our conclusion in previous paragraph that the ASM used with Group 1 was more effective in motivating the PSTs in the peer discussions than the ASM used with Group 2. Furthermore, the ANOVA analysis indicated that the PSTs' evaluations were differing significantly in Items 4, 12, 14, and 21. In Item 4, although the means of the evaluations suggested overall satisfactions of the PSTs in both groups, the PSTs in Group 1 provided higher points suggesting their enjoyment from the PI method. Similarly, the PSTs in Group 1 provided higher points for Item 12 showing that they found allocated time for the PI method as sufficient. Moreover, in comparison to Group 2, the PSTs in Group 1 evaluated the PI method as useful and the level of peer discussions as sufficient. This finding confirmed that the PSTs in Group 1 found peer discussions to be effective in their understanding of the course topics.

We determined the PSTs' overall satisfactions from the learning environment and process using the "Course Evaluation Form." The PSTs' evaluations of the first three items suggested their overall satisfactions from the course instruction, course activities, and overall performance. Although their evaluation scores were very much similar for the first two items, in Item 3, more PSTs in Group 1 appeared to satisfy from their overall performance than the PSTs in Group 2. In Item 4 and Item 5, the PSTs reported their most liked and disliked features of the course and suggested changes for the features that they did not like. As presented in Table 11, many PSTs in both groups stated the PI as an effective method that facilitated their learning of the course topics by increasing their interest. In this study, we used the ASM to increase the PSTs' course attendance and encourage them for participating in group discussions. According to the ASMs, the PSTs in Group 1 and Group 2 received 30% and 10% of the score of a question, respectively, based on their group-mate's response. Therefore, 16 PSTs in Group 1 complained about the drop in their PI scores in the absence of group-mates and stated this situation as the most disliked feature of the course. This finding showed their motivation towards the attendance which eventually affected their learning of the course topics.

In this study, the PSTs were selected from a single university with 46 PSTs, which was a relatively small sample, we could not investigate the effects of moderating variables such as gender. Future studies can use larger samples, from different institutions and varying disciplines and levels of learning, to provide more detailed information on PSTs' academic achievements and to obtain more generalizable results. Moreover, they can investigate the effects of using mechanisms such as the ASM in lessons on students' participation in group discussions, satisfaction from learning activities, and their academic achievement.

Acknowledgements Parts of this study were presented at the 2017 International Congress on Social Sciences, Madrid/Spain.

Compliance with ethical standards

Conflict of interest We have no conflicts of interest to disclose.

References

Adawi, T., Burden, H., Olsson, D., & Mattiasson, R. (2016). Characterizing software engineering students' discussions during peer instruction: Opportunities for learning and implications for teaching. *International Journal of Engineering Education, 32*(2), 927–936.

Arteaga, I. L., & Vinken, E. (2013). Example of good practice of a learning environment with a classroom response system in a mechanical engineering bachelor course. *European Journal of Engineering Education, 38*(6), 652–660.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. J. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education, 62*, 102–110.

Bruff, D. (2009). *Teaching with classroom response systems: Creating active learning environments*. San Francisco: Jossey-Bass.

Bulut, S., Ekici, C., & İşeri, İ. A. (1999). Bazı olasılık kavramlarının öğretimi için çalışma yapraklarının geliştirilmesi [The Development of worksheets for teaching some probability concepts]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 15*, 129–136.

Chou, C. Y., & Lin, P. H. (2015). Promoting discussion in peer instruction: Discussion partner assignment and accountability scoring mechanisms. *British Journal of Educational Technology, 46*(4), 839–847.

Cortright, R. N., Collins, H. L., & DiCarlo, S. E. (2005). Peer instruction enhanced meaningful learning: Ability to solve novel problems. *Advances in Physiology Education, 29*(2), 107–111.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics, 1*(1), 40–95.

Celik, D., & Gunes, G. (2007). 7 8 ve 9. sınıf öğrencilerinin olasılık ile ilgili anlama ve kavram yanılgılarının incelenmesi [Investigating 7th, 8th, and 9th grade students' understanding of probability and misconceptions]. *Milli Eğitim Dergisi, 173*, 361–375.

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics, 78*(10), 1056–1063.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*(2), 81–94.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher, 40*(4), 206–209.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2006). Peer instruction enhanced student performance on qualitative problem-solving questions. *Advances in Physiology Education, 30*(4), 168–173.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education, 53*(3), 819–827.

Latulippe, J. (2016). Clickers, ipad, and lecture capture in one semester: My teaching transformation. *Problems, Resources, and Issues in Mathematics Undergraduate Studies, 26*(6), 603–617.

Lee, C. B., Garcia, S., & Porter, L. (2013). Can peer instruction be effective in upper-division computer science courses? *ACM Transactions on Computing Education, 13*(3), 1–22.

Len, P. M. (2007). Different reward structures to motivate student interaction with electronic response systems in astronomy. *Astronomy Education Review, 5*(2), 5–15.

Lucas, A. (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus, 19*(3), 219–231.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River: Prentice Hall.

MacArthur, J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice, 9*(3), 187–195.

Michinov, N., Morice, J., & Ferrières, V. (2015). A step further in peer instruction: Using the Stepladder technique to improve learning. *Computers & Education, 91*, 1–13.

Miller, T. K. (2013). Pre-service elementary teachers' perceptions of a modified peer instruction implementation of clickers in their mathematics content course. *Issues in the Undergraduate Mathematics Preparation of School Teachers, 3*, 1–15.

Morice, J., Michinov, N., Delaval, M., Sideridou, A., & Ferrières, V. (2015). Comparing the effectiveness of peer instruction to individual learning during a chromatography course. *Journal of Computer Assisted Learning, 31*(6), 722–733.

Olpak, Y. Z., Karaoglan Yilmaz, F. G., & Yilmaz, R. (2017). Development of a student evaluation form toward peer instruction. *Turkish Online Journal of Educational Technology*, (Special Issue for INTE 2017), 839–845.

Rao, S. P., & DiCarlo, S. E. (2000). Peer instruction improves performance on quizzes. *Advances in Physiology Education, 24*(1), 51–55.

The International Society for Technology in Education (ISTE). (2017). *ISTE standards for students*. Retrieved from: https://www.iste.org/standards/for-students. Accessed 1 February 2018.

Trout, M. J., Borges, N., & Koles, P. (2014). Modified peer instruction improves examination scores in pharmacology. *Medical Education, 48*(11), 1112–1113.

Yourstone, S. A., Kraye, H. S., & Albaum, G. (2008). Classroom questioning with immediate electronic response: Do clickers improve learning? *Decision Sciences Journal of Innovative Education, 6*(1), 75–88.