International Journal of Progressive Education, Volume 18 Number 3, 2022

The Impact of Peer Review on Pre-Service Science Teachers' Written Arguments about Socioscientific Issues Related to Chemistry

Esra Capkinoglu [i]
Independent Researcher

Gulsen Leblebicioglu [ii]
Bolu Abant Izzet Baysal University

Duygu Metin Peten [iii]
Ege University

Pinar Seda Cetin [iv]
Bolu Abant Izzet Baysal University

Abstract

This study investigated the impact of peer review on developing pre-service science teachers' (PSTs) written arguments about socioscientific issues related to chemistry. In the study, a quasi-experimental design was used with experimental (32 PSTs) and comparison class (33 PSTs). The participants were PSTs who were juniors in a public university in Turkey. Argumentation procedure for each group was conducted by the same instructor identically except peer review of written arguments in the experimental class. Toulmin's (1958) argument pattern was used for coding the arguments, and argumentation levels were determined by the levels proposed by Venville and Dawson (2010). Results indicated that the experimental class generated more Level 3 and Level 4 argumentation comprises more complex arguments than the comparison class. In both groups, more complex arguments were generated in the contexts of the use of medicine and home chemicals whereas less complex arguments were generated in the context of chemical additives in food. The conclusions and implications for science educators and researchers were discussed.

Keywords: Peer Feedback; Peer Review; Pre-Service Science Teachers; Socioscientific Issues in Chemistry; Written Argument.

DOI: 10.29329/ijpe.2022.439.17

----

[i] **Esra Capkinoglu,** Dr., Elementary Science Education, Independent Researcher, ORCID: 0000-0003-3076-1196

Correspondence: yardimciesra@yahoo.com

[ii] **Gulsen Leblebicioglu,** Prof. Dr., Department of Mathematics and Science Education, Bolu Abant Izzet Baysal University

[iii] **Duygu Metin Peten,** Dr., Department of Mathematics and Science Education, Ege University

[iv] **Pinar Seda Cetin,** Assoc. Prof. Dr., Department of Mathematics and Science Education, Bolu Abant Izzet Baysal University

INTRODUCTION

Argumentation has been basically defined as a way of knowing in which students talk and write in the language of science (Duschl & Osborne, 2002; Sampson & Walker, 2012). Because of its interactive process that gives students the opportunity to express themselves in oral or written forms, students have a chance to participate and contribute to the solution of societal and controversial issues. However, they need to learn the requirements of argumentation to construct supported and warranted claims, conclude meaning, and write argumentative essays while considering for opposing viewpoints on the need of the controversial issues (Noroozi et al., 2017). Students would learn these processes from each other as well as their teachers. From this perspective, argumentative peer review and feedback can be seen as a part of the interactive process between learners. Peer feedback has been considered as a powerful instructional practice that teachers use to enhance students' writing skills, their motivation (Brown, 2005), and also their learning performances in the specific content domain (Nelson & Schunn, 2008). It was asserted in this study that peer feedback would also develop argumentative skills. In order to study this assertion, the PSTs participated in the various classroom discussions about socioscientific issues related to chemistry and constructed their own written arguments.

The idea of peer review of written arguments was inspired by Argument Driven Inquiry (ADI) which is an instructional model for helping students learn how to participate in scientific argumentation and craft written arguments (Sampson et al., 2011). In one of the eight steps of ADI, the students review investigation reports of their peers like a referee. This step provides an opportunity for students to get developmental feedback from their peers on the quality of their work. It also encourages students to understand appropriate standards for "what counts" as quality and to develop metacognitive skills. Thus, they would determine the validity or the acceptability of a claim or evidence. Furthermore, they have a chance to distinguish strong arguments from the weak ones. The present study, therefore, is designed by considering these features of ADI in order to determine the impact of peer review on PSTs' written arguments about some socioscientific issues related with chemistry.

In literature, there have been some research studies about the effectiveness of ADI instructional model in developing students' argumentative writing skills especially in scientific contexts (Cetin & Eymur, 2017; Metin Peten, 2019; Sampson et al., 2013; Sampson et al., 2011; Sampson & Walker, 2012; Walker & Sampson, 2013). However, there have been very limited studies combining the ADI model and SSI contexts, provided that it is limited to available resources. Grooms et al. (2014) investigated whether or not undergraduate students who craft arguments related to SSI-based context associated with the environment and personal health before and after completing an argument-based chemistry laboratory course and traditional chemistry laboratory course. According to the researchers, the ADI treatment group consisted of rationales in their arguments more often than the group experiencing the traditional chemistry instruction. ADI was not directly used as an instructional strategy in the present study, but it is believed that the results obtained from this study will contribute and gap the fill in socioscientific argumentation literature in order to show the effect of peer review (as in ADI model) on PSTs' written argumentation in socioscientific contexts.

The study was informed by ADI, peer review, and argumentation literature which were summarized in the following sections.

Theoretical framework

Argument Driven Inquiry (ADI)

ADI is an instructional model (Walker & Sampson, 2013) that provides an opportunity for students to learn how to develop a methodology to solve research questions, to collect data, to use data to answer a research question, write, and be more reflective as they work (Sampson et al., 2011) in science laboratories. Additionally, it ensures a comprehensive scientific argumentation and peer review process during laboratories (Sampson et al., 2011). The ADI instructional model includes eight stages that are intended to provide a template or guide for designing students' laboratory experiences to promote scientific argumentation (Grooms et al., 2014). These stages are identifying and guiding the research question (1), designing a method and collecting data (2), analysing data and developing a tentative argument (3), argumentation (4), explicit and reflective discussion (5), writing a research report (6), peer review (7), and revising and submitting the report (8).

The first step of ADI involves presenting students with a topic and providing a research question for their answer by the teacher. In the second step, each group develops a method to collect data, carries out their research plan, and collects data. In the third step, each group analyses their data and generates an argument consisted of a claim, evidence, and justification. They also prepare their presentations such as a whiteboard or a poster that can easily be seen by others. In the fourth step, each group shares their argument and others ask questions and offer critiques. In the fifth step, the students discuss what they learnt about the scientific knowledge and nature of science under the direction of teacher. In the sixth step, each student write a report consisted of three sections: the goal of the investigation, the method that they used, and their argument. In the seventh step, the reports are evaluated by considering a peer-review guide and feedback is given. In the last step, each student revises their report based on the results of the peer-review process and submits to their teacher for a final evaluation (Grooms et al., 2014).

In the related literature, there are different studies investigated the role of the ADI instructional model on different variables. For instance, some researchers found that the students produced better arguments after they participated ADI based laboratory activities (Grooms et al., 2014; Sampson et al., 2011; Walker & Sampson, 2013). Some other studies that used ADI with explicit-reflective NOSI indicated that this approach is positively effective for promote students' understanding about NOSI (Metin Peten, 2021; Eymur, 2019).

As stated before, the ADI instructional model was not used as an instructional strategy in the present study, but it was thought that it would be useful to give some literature on the ADI.

Peer review and teacher education

In the literature of assessment for learning, peer review, which is also used as "peer assessment", "peer evaluation" or "peer feedback", has been acknowledged as an indispensable technique for writing instruction (Breuch, 2004). Falchikov (2001) describes peer review as students involving "in reflective criticism of the work or performance of other students using previously defined criteria and supply[ing] feedback to them" (pp. 2–3). Receiving feedback from peers who are learning partners with the same motivational requirements and also giving feedback to them in a reciprocal manner are crucial aspects of learning process (Bayerlein, 2014). Throughout this process, students are actively included in the learning of how to learn (Tillema et al., 2011). Thus, the peers, whose social status is the same, are cut down immediate teacher intervention and allowed them to learn with and from each other (Boud et al.,1999).

Since there are some limitations of the teacher assessment of students understanding, students' educational goals, and how these connect with educational applications (McLaughlin & Simpson, 2004), peer review can add on it in various disciplines and in various education levels. More specifically, the implementation of peer review in an educational setting has some advantages such as the opportunity to think and understand comprehensively when the students compare their own writing processes and products with those of others (Yang, 2010). Besides, it develops many social and communication skills such as verbal lecturing skills as well as the abilities of criticizing and openness to criticism (Topping et al., 2000), and encourages students to become self-regulated learners (Carless, 2006).

Peer review is also an effective method in developing teaching skills in teacher education (Bagci Kilic & Cakan, 2007; Sluijsmans & Prins, 2006). These researchers clarify the importance of peer review in teacher education with three reasons. First, due to the students' cooperation and communication with each other in peer review process, the student teachers would improve these skills and become members of an organization when they become teachers. Second, their reflection skills that are necessary for making reliable judgments during peer works would strengthen through this method. Third, the student teachers learn how they make critical judgments about their peers' performances, and thus they would make better critical judgments about students' work in their own classes.

There are a few studies of peer review with pre-service teachers for developing their teaching skills. For instance, a research conducted with prospective teachers investigated how peer review contributed to the students' development in their field experience. The results showed that it has a positive influence on promoting some educational competencies such as forming standards for reflection and assessment, displaying self-confidence and attitudes to peer review (al-Barakat & al-Hassan, 2009). Another study conducted with prospective elementary teachers examined their opinions about peer assessment of their teaching practice. The students declared that peer assessment had a positive impact on their assessment skills and improved their awareness of their strengths and weaknesses (Koc, 2011). It was also evidenced that pre-service teachers could assess their peers' science teaching performance reliably (Bagci Kilic & Cakan, 2007). Although peer scores were considerably higher than the instructor scores, they were significantly correlated with the instructor scores.

While the importance of peer review in the literature is emphasized in aforementioned studies, the following section relates peer review to argumentation.

Relationship between argumentation and peer review

Sanmartí (2003) defines argumentation as follows;

> "A social, intellectual, and verbal activity that serves to justify or refute an opinion, and consists of making statements taking into account the recipient and the purpose of communication. To argue is then to choose from different options or explanations and to reason the criteria for assessing why the chosen option was the best." (p. 123)

In this sense, the process of arguing is to construct a contention by providing persuasive justifications based on strong evidence. In other words, there are some essential structures of an argument that make the argument distinctive. For determining the structure of an argument, Toulmin's model (Toulmin, 1958) is generally used in science education. According to this model, the components of an argument are "claim", "data", "warrant", "backing", "qualifier", and "rebuttal". The claim is the opinion or the conclusion on a specific issue; the data is the evidence to support that claim; the warrant is the justification shows the connection between the claim and the data; the backing is the theoretical assumptions supported warrant; the qualifier shows the limits of the claim; and rebuttal is the circumstances under which the claim is invalid. Argumentation practices based on these components progress in two ways which are oral (classroom discourse) and written (students' reports or online text) argumentation in science education. In the present study, the PSTs involved in classroom discussions and wrote their individual arguments. Since they reviewed each other's written arguments according to aim of the present study, just their written arguments were included as data source.

In the literature, writing argument in socioscientific issues is a common practice for students, for example in the field of genetics (Venville & Dawson, 2010; Zohar & Nemet, 2002), nuclear power plants (Demircioglu & Ucar, 2014), biodiversity (Evagorou & Osborne, 2013), climate change (McDonald, 2014), and a local socioscientific issue (Kolsto, 2006). Many of these studies reported that students often lack of constructing quality argumentation. Two reasons for this might be that they are unaware of the features of high quality argumentation writing or they have difficulty in transferring their knowledge into the practices (Bacha, 2010). In this respect, different researchers suggest that the use of peer review (and peer feedback), which can be considered as a powerful vehicle to facilitate writing argument (Hattie & Timperley, 2007), and understand and internalize the criteria for a quality argument (Osborne et al., 2016). Thus, in peer review process, students are provided with the opportunity for critically testing and considering the peer's perspective and counter-arguments, and also support their own perspective by taking the peer's perspective into account (Noroozi et al., 2017; Weinberger & Fischer, 2006).

Argumentative peer review has also been conducted with undergraduate and high school students. Two studies investigated the effects of online peer feedback on undergraduate students' writing argumentative essays on Genetically Modified Organisms (Noroozi & Hatami, 2018) and Life Sciences (Haro et al., 2018). Another study conducted by Lu and Zang (2013) investigated whether using an online rubric to assess the written arguments of peers can lead high school students to write better arguments.

The authors of this study also claim that peer review would improve pre-service teachers' written arguments as it was the case in developing their science teaching skills (Bagci Kilic & Cakan, 2007; Koc, 2011). In addition, if peer reviews were shared immediately in a class session and used in developing their first draft argumentations, it would be even more effective. Indeed, it is important to know that it is not always practical for an instructor to provide immediate feedback on each student's argument because of class size, lack of time, and additional effort etc. However, peers would recognize strong and weak parts of an argument while reviewing their partner's argument through critical thinking and making sophisticated judgments in the light of some criteria. This also would help them in improving their own written argument on return. These are all common sense claims and need further inquiry to see if it really makes a difference.,

From this stand, the current study was aimed at determining the effectiveness of peer review on developing the PSTs' writing arguments on socio-scientific issues related to chemistry. Thus, the specific research question guiding this study is "How peer review of PSTs' arguments in dyads affects their written arguments on socioscientific issues related to chemistry?".

METHOD

In order to determine the effect of peer review on pre-service teachers' arguments, a quasi-experimental design was applied. The study was conducted in two classes of a course, Special Topics in Chemistry, in a public university in Turkey. The course content was chemistry concepts in social and daily life such as Climate change & Kyoto Protocol, chemical additives in food, use of medicine, home chemicals, and chemical pollution & acid rain. The aim of the course was to teach the course content and help PSTs developing their own perspective on these concepts. In order to make PSTs think more and clear out their ideas about these concepts, the researchers decided to require them to write an argument just after discussion of each topic in the class. In order to see how adding peer review into their argumentation process affects PSTs' argumentation skills, pre-service teachers' written argumentations in the experimental class were reviewed in dyads, returned to the owner, and let them develop their arguments based on peer review. After getting feedback from their peers, the PSTs studied to develop their written argument and wrote their arguments at the end of the class. Finally, they submitted the final form of their arguments at the end of the class session. In order to determine the effect of peer review on argumentative skills, the same applications were conducted in the comparison class except peer review. The PSTs in the comparison class constructed their written arguments and submitted the initial version of it at the end of the class session. These applications were repeated in the same way on five socioscientific issues related to chemistry. All other factors such as instructor, course topic, discussions were the same in both groups. The implementation of the study was explained in detail in implementation section after defining the participants in the following session.

Participants

In total, 65 PSTs enrolled in the study. Thirty-two of them were enrolled in the experimental class whereas 33 of them were in the comparison class. The PSTs' ages ranged from 19 to 21 years. They were at the beginning of their third (junior) year of the Elementary Science Teacher Education Program. There are both pure science and educational pedagogy courses in previous two years of the program. In the first year, they had taken various science courses such as General Chemistry, General Physics, and Mathematics Courses. At the second year, they had taken more specific science courses such as Organic Chemistry, Analytical Chemistry, Optics, and Modern Physics in addition to two educational pedagogy courses named Instructional Principles and Methods, and Science and Technology Curriculum. The course named Special Topics in Chemistry was one of the compulsory courses in the fall semester at the third year.

Implementation procedure

The course instructor was a professor (second author), and she had two classes of the same course. One of the classes was randomly assigned as experimental class and other group was considered as a comparison class. The class time was two hours per week for each class. An illustration of the implementation procedure is presented in Table 1. At the beginning of the implementation, both of the classes were given argumentation training at the first two weeks, because argumentation was new for the PSTs. In this training, argumentation components specified by Toulmin (1958) were introduced and examples of arguments were shown. Then, PSTs were required to determine argument components in a scientific writing. After that, they were asked to write arguments on simple tasks such as choosing type of bottles for a drink and choosing baby sitter for their child. A few of them shared their arguments with class and argument elements in their arguments were made explicit for the class.

After argumentation training, two weeks were allocated for each of five issues which were Climate change & Kyoto Protocol, chemical additives in food, use of medicine, home chemicals, and chemical pollution & acid rain in both experimental and comparison class. Thus, the main implementation lasted ten weeks just after two weeks of argumentation training. The implementation procedure for each class was the same except peer evaluation of written arguments in the experimental class.

Table 1. The illustration of implementation procedure in the experimental and comparison class.

Argumentation Training (First two weeks)	First week of each issue (Two class sessions, each 40 min.)	Second week of each issue (Two class sessions, each 40 min.)
-Introduction of elements of an argument (Toulmin, 1958). -Determining argument elements in a scientific writing. -Writing two arguments on simple issues such as choice of bottle for a drink and choice of baby sitter for a child.	-Searching the issue on the Internet and bring information to the class. -Sharing the information in the class. -Deciding a problem related to the issue to discuss. -Determination of discussion groups. -Each group's taking a different perspective on the issue and searching more information to defend their perspective until the following week.	-Groups were given ten minutes to get prepared for discussion in front of the class. -Two members from each group come in front of the class. -They discuss the issue in a panel like setting. -The instructor open up, monitors and close the discussion. -Writing individual arguments. -Peer review of individual written arguments (only in experimental class). -Returning peer reviews to the partner and improving their arguments based on peer review (only in experimental class).

Before the first week of each issue, readings on the Internet were announced in an online platform. These readings were made from Google, official websites, online newspapers, etc. The PSTs were asked to read those readings and were suggested to do additional search on the Internet related to the subject under investigation. The instructor never lectured during the course. Instead, she asked the

PSTs to share their knowledge and understanding of the issues and tried to clarify different views on the issue. There were always different perspectives on an issue.

Then, the instructor had asked the class to decide on a problem to discuss in the following week. After that, different perspectives on the issue were also determined. Thus, first week of each issue was ended with a clear problem to discuss in the following week and groups of PSTs who will take position on different (sometimes opposite) side of the issue were determined. Groups of PSTs were required to search more knowledge until next week to support their perspectives and get ready for discussion between two opposing groups in the following week.

In the second week of each issue, the discussion was held between two groups of PSTs approximately 30 min and a jury was assigned to observe and determine the winner of the discussion. Since class size was large, groups of PSTs on each side were around ten or higher. Additionally, since argumentation between large groups of the PSTs would be difficult, each opposing group was asked to select five members from their group to present and defend their perspective in the discussion in a panel format in front of the classroom. Thus, there were five PSTs in each side and a jury of five PSTs. The rest of the class observed the discussion.

The instructor started the discussion, monitored it if it got out of the issue. She also suggested asking questions to each other, answering by justifying the PSTs' answer by evidence from their knowledge search. Each member was given five minutes to present their group's perspectives, to explain their argument on the issue, and then both groups questioned each other's claims, asked questions, and asked for evidence, etc. Other students listened to the discussion and asked questions. In order to include more student decision into class, a jury consisting of five students was chosen from the PSTs that were supposed to listen the discussion. At the end of the discussion, the jury determined which group discussed the issue better than the other group. The jury changed each week. After a short recess, each PST wrote their individual arguments on the issue in the class on a worksheet that was provided by the instructor.

The format of the worksheets was open-ended. The PSTs were asked to explain their claims based on the discussion and write their comprehensive arguments which included supportive elements. Specifically, they were asked to write an argumentation on the following questions in each of the issues respectively:

1. Climate change & Kyoto Protocol: What is your idea about our country signing the Kyoto Protocol? Is it reasonable for our country or not? Why?

2. Chemical additives in food: What is your idea about adding chemicals into our foods for various purposes such as preserving, sweetening, etc.?

3. Use of medicine: What is your idea about herbal or chemical drugs? Which one do you prefer and why?

4. Home chemicals: What is your idea about organic and chemical detergents? Which one do you prefer and why?

5. Chemical pollution & acid rain: Do you think if industrial organizations should take urgent precautions in order to prevent environmental pollution or not?

The PSTs didn't have to claim and defend the same idea that they got prepared and discussed in the class. They were allowed to change their opinion on the issue which means that they would defend the opposite idea of the one that they defended during the discussion. This two-week procedure was repeated in the same manner in both experimental and comparison class for each issue. Thus, each participant constructed five arguments on five aforementioned issues. The difference in experimental class is explained more detailed in the next section.

Difference in experimental class

The only difference between experimental and comparison class was peer review of individually written arguments in the end. In the comparison class, the PSTs handed in their arguments to the instructor at the end of the class. On the other hand, in experimental class, after finishing their arguments, the PSTs were paired with another PST in dyads and exchanged their arguments. They were asked to evaluate their partner's arguments on a set of criteria stated on a peer review form provided by the instructor.

Peer review form was asking if the argument has a clear claim, evidence for the claim, backed with data, and have a qualifier (presented in Appendix-1). Indeed, it directly asked if Toulmin elements exist in the argument under investigation in order to evaluate the quality of the argument. Peer review form was easy to use format with yes or no check for each question and organized around five questions:

- Did the author explain his/her claim explicitly?
- Did the author use data or evidence to support his/her claim?
- Did the author justify enough the claim when generating his/her argument?
- Did the author support his/her warrants appropriately?
- Did the author specify in which conditions his/her claim was valid?

In addition to these questions, there was an open-ended section asking for feedback for improving the quality of the partner's argument. After completing their review, the PSTs returned the argument to their partner and get their peer reviewed argument. They were suggested to make developments on their arguments if they needed to. They mostly worked on developing some parts of their argument rather than re-writing it. They handed in last version of their written arguments at the end of the class.

Data analysis

The PSTs' written arguments were first coded according to Toulmin's (1958) argument pattern components such as claim, data, warrant, backing, and qualifier. After coding process, each PST's argumentation level was determined by a coding scheme presented in Venville and Dawson (2010). This coding scheme was chosen, because it is in accordance with Toulmin framework and clarifies the complexity of arguments in an understandable and easy to follow format. These researchers categorized arguments into four levels based on the presence of argument elements. Each level of argumentation was defined below.

- Level 1: Claim only (statement, conclusion, proposition),
- Level 2: Claim, data (evidence supporting the claim) and/or warrant (relationship between claim and data),
- Level 3: Claim, data/warrant, backing (assumptions to support warrant) or qualifier (conditions under which claims are true),
- Level 4: Claim, data/warrant, backing, and qualifier (Venville and Dawson 2010).

In order to present the coding scheme in detail, an example for each level of argumentation in use of medicine issue is presented below.

> ***Level 2 argumentation:*** We should prefer the use of herbal medicine [Claim]. Because 250.000 people die due to the medical errors in USA in every year. 127.000 of these people die because of wrong medicine or side effects of drugs when they are treated in hospital [Data]. Chemical drugs play a major role in these deaths by direct destructive effects to immune system, blood circulation, and reproductive system. For example, some drugs demolish the blood cells by fusing their walls. If the patient seriously ill, his/her blood production deteriorates and leukemia or anemia occurs due to the suppression of bone marrow [Warrant]. Conversely, herbal medicines are produced from more natural active ingredients [Data]. Instead of chemical drugs, herbal medicines such as stinging nettle, rosehip, chamomile, turmeric, apple, or lavender are the least damaging medicine for cells [Warrant]. (PST 1, comparison class).

In this written argument from comparison class, the PST presents a clear claim that supported with evidence. The PST also provides the connection between data and claim explicitly. Since this argument comprises the components of claim, data, and warrant, it is an appropriate example for level 2 argumentation.

> ***Level 3 argumentation:*** Use of chemical drugs is more reliable than herbal medicines [Claim]. Because, chemical drugs are made of certain doses of substances and produced by certain protocols. They treat diseases in a short time. They also contain certain proportions of herbal ingredients. But, although the use of herbal medicines would be healthy, it is not reliable because there is not an institution to control them. People take them on their own decision. There is no certificate on how herbal medicines are prepared and which plants are used in its content. On the other hand, there is a prospectus in chemical drugs defining which substances are found, which dose should be used, side effects, and duration of treatment. By this way, people have information about the drug that they use. Since chemical drugs are prescribed in a controlled manner by experts, they are less harmful to human health [Data+Warrant]. My argument is valid in the circumstances that a patient consults a doctor for a treatment and takes drugs as prescribed. In other cases, overdose of medicine would have a harmful effect on human health [Qualifier]. (PST 2, experimental class).

The argument above from a student in the experimental class. The PST backs up his/her claim with data and warrant components appropriately. Besides, the PST specifies the conditions under which his/her claim is true. Since this written argumentation includes the components of claim, data, warrant, and qualifier, it is an example of Level 3 argumentation.

> ***Level 4 argumentation:*** Herbal medicine should be used [Claim]. Phytotherapy means treatment with plants. People have benefited from plants since ancient times. The biggest difference between past and present of phytotherapy is that only beneficial parts of the plants are used for treatment not the whole as in the past. This keeps the patient away from other parts of the plant, which may have side effects [Data+Warrant]. Herbal products are mentioned in the book of the Hippocrates who is considered as the father of medicine in ancient times. Similarly, Ibn-i Sina and Al Gafini have important works on herbal medicine. In recent years, herbal treatment and their specialists have significantly increased [Data]. The official newspaper on May 27, 2004 has also approved the sale of herbal medicines in pharmacies and auditing of them [Backing]. But we must use herbal medicines consciously under the supervision of the expert, not unconsciously [Qualifier 1]. Because the herbal products which are natural do not mean that it does not harm the user. Toxicological effects of plants or herbal products when used in combination with certain drugs have been determined. Therefore, it is important to use them under the expert supervision. It has been said that there is no disease that cannot be healed after using appropriate dosage of herbal medicines determined by the experts [Data+Warrant]. However, this also has limitations. Yes, it is treated, but the duration of treatment is longer than the treatment with chemical medicines [Qualifier 2]. Thus, you need to be patient. (PTS 3, experimental class).

In this last written argument from experimental class, the PST asserts his/her claim with various data and provides the link between these data and the claim. The PST also supports his/her justification with a backing component that contains a declaration made by official newspaper. Moreover, he/she specifies the boundaries of his/her argument by stating the conditions with data and justifications. Since the student used all of the components properly and constructed a well-structured argument, this argument is categorized as Level 4 argumentation.

Reliability of coding

Three researchers who were the co-authors of the present study coded the data. They applied the coding scheme on randomly selected five arguments (one argument from each socioscientific context) together to establish a shared understanding of the determination of argument elements in the arguments. Then, they independently coded other randomly selected five arguments and compared their codes. The three coders were blind to the condition (i.e. experimental or comparison class) for each argument. They first determined the argument elements in an argument and then decided which level it was. They counted agreements between each coder and divided by total codes as stated in Miles and Huberman (1994). There were 15 possible agreements in total and there was only one disagreement. This disagreement was counted two, because this coder disagreed with other two coders. Thus, number of agreements was accepted as 13 out of 15, and inter-rater agreement was calculated to be 87%.

After that, the researchers shared the PSTs' arguments in both of the classes issue by issue and coded the data independently in a day. Thus, PSTs' arguments on an issue in experimental and comparison class were coded by the same researcher. The researchers met the day after. They took an additional action to increase reliability of the coding. They randomly selected five PST arguments that they coded and exchanged their coded data with each other. They did not put any notes on the data. They coded the exchanged data independently from each other. In this way, fifteen PST arguments were coded by all three coders. They compared their codes for each PST and counted the agreements and disagreements between the coders. There were 12 agreements in 15 possible agreements and coding agreement was 80%.

They shared the rest of the PSTs' arguments equally again in the same way as issue by issue and coded in a day. They met the following day and repeated the procedure just explained in the previous paragraph. They agreed in all of the codes at this time and coding agreement was 100%. The frequencies of the argumentation levels were organized and their percentages presented in figures including graphs to visualize the data.

RESULTS

The percentages of the argumentation levels on five issues are presented in the following figures for experimental and comparison class. The classes were compared level by level on each socioscientific issue. Figure 1 represents the percentages of argumentation levels generated by the classes in the context of Climate change & Kyoto Protocol. None of the groups generated Level 1 argumentation. While the comparison class generated more Level 2 argumentation than the experimental class, the experimental class generated more Level 3 argumentation than the comparison class. Only 4% of experimental class and only 3% of comparison class generated Level 4 argumentation. Another finding is that the total percentage of Level 3 and Level 4 argumentations which represent more sophisticated argumentations is higher in the experimental class.

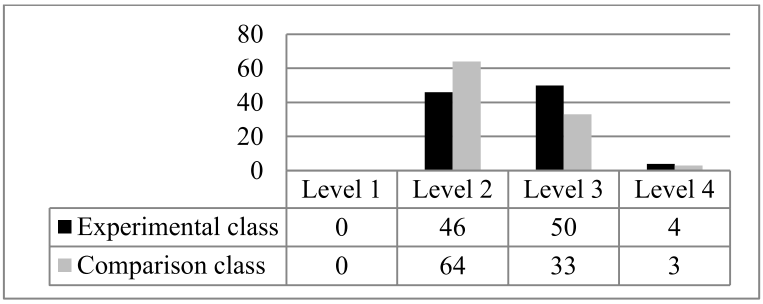


Figure 1. Percentages of argumentation levels generated by the classes in the context of Climate change & Kyoto Protocol.

The percentages of the argumentation levels generated by the classes in the context of chemical additives in food are presented in Figure 2. According to Figure 2, the classes didn't generate any Level 1 argumentation. The comparison class generated more argumentation in both Level 2 and Level 3 compared to the experimental class. However, the PSTs in this class could generate any Level 4 argumentation whereas 21% of the argumentation is belonged to this level in the experimental class. As in the Climate change & Kyoto Protocol context, the percentages of more sophisticated argumentation (Level 3 and Level 4) are high in the experimental class.

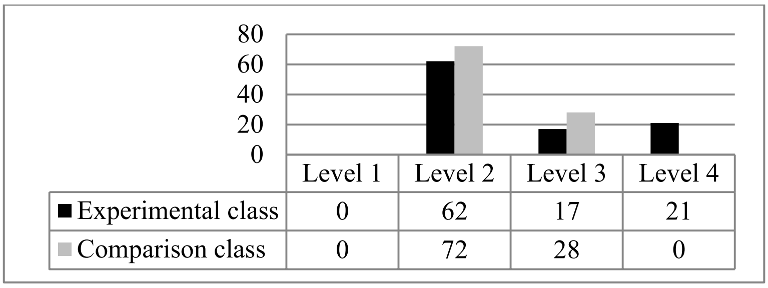


Figure 2. Percentages of argumentation levels generated by the classes in the context of chemical additives in food.

Figure 3 represents the percentages of the argumentation levels generated by the classes in the context of use of medicine. In this context, any Level 1 argumentation generated by the classes. The comparison class generated more Level 2 argumentation than the experimental class. Conversely, the experimental class generated more Level 3 argumentation than the comparison group. 17% of the students in both classes generated Level 4 argumentation. But, if the number of more sophisticated arguments were compared, the experimental class produced more than the comparison group.

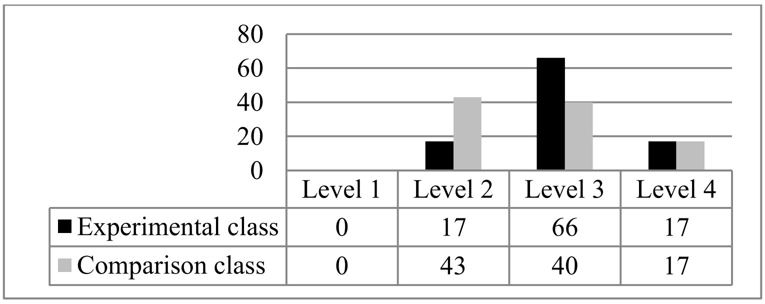
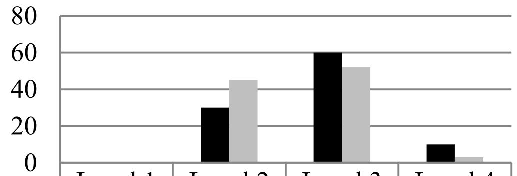


Figure 3. Percentages of argumentation levels generated by the classes in the context of use of medicine.

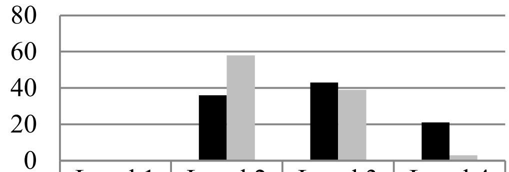
As it is shown clearly in Figure 4, Level 1 argumentation wasn't produced in both classes in the context of home chemicals. The comparison class generated more Level 2 argumentation whereas the experimental class generated more Level 3 argumentation. For Level 4 argumentation, 10% of the experimental class and only 3% of the comparison class generated this argumentation level. The experimental class again generated more sophisticated argumentations in this context.



	Level 1	Level 2	Level 3	Level 4
■ Experimental class	0	30	60	10
■ Comparison class	0	45	52	3

Figure 4. Percentages of argumentation levels generated by the classes in the context of home chemicals.

Lastly, the percentages of the argumentation levels generated by the classes in the context of chemical pollution & acid rain are presented in Figure 5. Both classes generated any Level 1 argumentation as it was the case in previous contexts. The comparison class generated more argumentation in Level 2 whereas the experimental class generated more argumentation in Level 3. On the other hand, the experimental class generated more Level 4 argumentation than the comparison class. Additionally, the PSTs in this class generated more sophisticated argumentations compared to the PSTs in the comparison class.



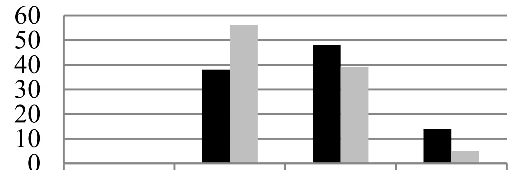
	Level 1	Level 2	Level 3	Level 4
■ Experimental class	0	36	43	21
■ Comparison class	0	58	39	3

Figure 5. Percentages of argumentation levels generated by the classes in the context of chemical pollution & acid rain.

When comparing each issue, the comparison class generated more Level 2 argumentation than the experimental class in all of the contexts. On the other hand, the experimental class generated more Level 3 argumentation than the comparison group in four contexts which are Climate change & Kyoto Protocol, use of medicine, home chemicals, and chemical pollution & acid rain. In addition, the experimental class had more sophisticated argumentations (total of Level 3 and 4) in all of the contexts.

Both classes generated highest number of sophisticated argumentations in two contexts which were the use of medicine (83% of the experimental class and 57% of the comparison class) and home chemicals (70% of the experimental class and 55% of the comparison class). On the other hand, both classes generated the lowest number of sophisticated argumentation in the context of chemical additives in food (38% of the experimental class and 28 % of the comparison class).

We also combined data for each issue and compared both classes (see Figure 6). While the comparison class generated Level 2 argumentations more than the experimental class, the experimental class generated more Level 3 argumentation. Moreover, the experimental class generated more sophisticated argumentations (Level 3 and Level 4) than the comparison class across the five contexts.



	Level 1	Level 2	Level 3	Level 4
Experimental class	0	38	48	14
Comparison class	0	56	39	5

Figure 6. Percentages of argumentation levels generated by the classes across all of the contexts.

DISCUSSION

This study examined how peer review of PSTs' arguments in dyads affected the quality of their written arguments about socioscientific issues related to chemistry. The findings indicated that the experimental class who reviewed each other's argument generated more Level 3 and Level 4 argumentation (62% of arguments in this class) than the comparison class in which 44% of arguments are at these levels. This difference in the number of higher level arguments indicates that peer review application in the experimental class developed their argumentation skills. During peer review, they would have been better recognized the criteria for a qualified argument than their counterparts in the comparison class. In addition, peer feedback provided in peer review in the experimental class would have helped them in improving their arguments.

A related study in the literature is about the comparison of the effect of peer and teacher review in the development of written arguments of PSTs. Metin Peten (2019) investigated which one of the effect of peer or teacher review would contribute more to the development of written arguments of PSTs in the review stage of ADI instructional model. In one of the comparison groups, she used ADI in its original form (peer review); while in the other group peer review was changed as lecturer review. In both groups, four ADI research activities were conducted. In order to see the difference between the groups, the PSTs were asked to write three different arguments apart from they studied in the class at the beginning and end of the semester. She found that the quality of the arguments created in each context in both groups increased from the beginning to the end of the semester as in the present study, but in terms of evidence and justification, argument-writing skills of PSTs reviewed by the lecturer improved better.

Similarly, Larson et al.'s (2009) study demonstrated that college and high school students who used rubrics to evaluate the structure of arguments and who received immediate feedback improved their ability to determine the quality of arguments. Other studies have also highlighted that peer review, where students can assess their peers' arguments, enables them to improve their arguments (see Haro et al., 2018; Lu & Zang, 2013; Noroozi et al., 2016; Noroozi & Hatami, 2018). According to the aforementioned studies in the related literature, it could be concluded that peer review improve the quality of students' and pre-service teachers' written arguments.

Although the experimental class showed better performance than the comparison class in generating complex arguments in all the contexts by the effect of peer review, the percentages of Level 4 argumentation were the least in both classes (14% of arguments in the experimental class, 5% of arguments in the comparison class). While the experimental class mostly generated Level 3 argumentation in total (48% of all arguments), the comparison class mostly generated Level 2 argumentation in total (56% of all arguments). This result shows that the PSTs have difficulty in constructing Level 4 arguments which have all argumentation elements. There have also been several studies showing that people from various education levels have difficulties in constructing a well-structured argument (Acar et al., 2010; Sadler, 2004). One possible reason for this result in the present study might be that poor writing skills of the students. According to our general observations, since multiple choice tests are commonly used in Turkish schools and all important high-stake exams such as high school and university entrance exams, our students from primary school to university are accustomed to choice one correct answer without explaining in written form. Because of these types of exams and limited writing practices in the classrooms, students often find it difficult to express themselves by writing. The PSTs in this study, therefore, might have difficulty in combining writing and constructing argument.

When the results are examined in terms of the issue, it could be said that the context of the issue was influential on generating quality arguments. More specifically, both experimental and comparison classes had the highest percentages of sophisticated argumentations (Level 3 and Level 4) in the contexts of the use of medicine and home chemicals. On the other hand, both classes had the lowest percentages of sophisticated argumentations in the context of chemical additives in food. This difference of the argumentation quality could stem from the content knowledge as well as oral argumentation made by the students during implementations. As it is evident in literature, there are many studies reported that the students' content knowledge is a significant factor in constructing qualified arguments (Dawson & Schibeci, 2003; Maloney & Simon, 2006: Sadler, 2004). Beside the content knowledge, classroom discussions might also affect the students' decisions. While the students act in accordance with the flow of the group in the oral discussions, they may have different performances during the writing argument individually in each issue.

Conclusion and suggestions for future research

The major finding of the present study demonstrated that using peer review in a sample of PSTs on socioscientific issues in Chemistry had positive impact on improving their written argumentation. Feedback in peer review forms would have been helped them in improving their arguments. Therefore, implementing peer review is useful for instructors who intend to provide students with timely high quality feedback, while at the same time intending to reduce the effort required to provide such feedback.

The second major finding of the present study demonstrated that both of the groups' argumentation levels changed in each issue. While the contexts of use of medicine and home chemicals were the most suitable contexts for promoting complex arguments, the context of chemical additives in food was the least suitable context for promoting sophisticated arguments in both experimental and comparison class. Since the nature of the contexts might influence the performance of the students in discussions, science educators should be careful in selection of the context. In addition, they would caution in judging the quality of student argumentations and consider the familiarity of the issue to the students.

This study only investigated the impact of peer review on developing the quality of PSTs written argument in five different socioscientific issues. It would be interesting to see how they perceive peer review or which of other feedback techniques could be used in addition to paper-based feedback. Since peer review research with PSTs was scarce, there is a need for similar research. We, therefore, suggest that further research could be investigate different variables on PSTs from different samples. Since, this study was limited with PSTs, similar studies would be held with different students in lower education levels for participating into argumentative peer feedback practices in the classrooms.

Disclosure statement

No potential conflict of interest was reported by the authors.

REFERENCES

Acar, O., Turkmen, L., & Roychoudhury, A. (2010). Student difficulties in socio-scientific argumentation and decision-making research findings: Crossing the borders of two research lines. *International Journal of Science Education, 32*(9), 1191–1206.

al-Barakat, A. A., & al-Hassan, O. A. (2009). Peer assessment as a learning tool for enhancing student teachers' preparation. *Asia-Pacific Journal of Teacher Education*, *37*, 399–413. https://doi.org/10.1080/13598660903247676

Cetin, P. S. & Eymur, G. (2017). Developing students' scientific writing and presentation skills through argument driven inquiry: An exploratory study. *Journal of Chemical Education*, *94*(7), 837- 843. https://doi.org/10.1021/acs.jchemed.6b00915

Metin Peten, D. (2019). Comparison of the effect of peer and teacher assessment in the development of written arguments of pre-service teachers. *Journal of Education for Life, 33*(2), 121–139. https://doi.org/10.33308/26674874.2019332108

Metin Peten, D. (2021). Influence of the Argument-Driven Inquiry with explicit-reflective nature of scientific inquiry intervention on pre-service science teachers' understandings about the nature of scientific inquiry. *International Journal of Science and Mathematics Education,* https://doi.org/10.1007/s10763-021-10197-8

Bacha, N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes*, *9*(3), 229–241. https://doi.org/10.1016/j.jeap.2010.05.001

Bagci Kilic, G., & Cakan, M. (2007). Peer assessment of elementary science teaching skills. *Journal of Science Teacher Education*, *18*(1), 91–107. https://www.jstor.org/stable/43156408

Bayerlein, L. (2014). Students' feedback preferences: How do students react to timely and automatically generated assessment feedback?. *Assessment and Evaluation in Higher Education*, *39*(8), 916–931. https://doi.org/10.1080/02602938.2013.870531

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, *24*(4), 413–126. https://doi.org/10.1080/0260293990240405

Breuch, L. K. (2004). *Virtual peer review: Teaching and learning about writing in the online environments*. SUNY Press.

Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education*, *1*(1), 81–89.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, *31*, 219–33. https://doi.org/10.1080/03075070600572132

Dawson, V., & Schibeci, R. (2003). Western Australian high school students' attitudes towards biotechnology processes. *Journal of Biological Education*, *38*(1), 1–6. https://doi.org/10.1080/00219266.2003.9655889

Demircioglu, T., & Ucar, S. (2014). Investigation of written arguments about Akkuyu Nuclear Power Plant. *Elementary Education Online*, *13*(4), 1373–1386. https://doi.org/10.17051/io.2014.31390

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72. https://doi.org/10.1080/03057260208560187

Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. *Journal of Research in Science Teaching*, *50*(2), 209–237. https://doi.org/10.1002/tea.21076

Eymur, G. (2019). The influence of the explicit nature of science instruction embedded in the argument-driven inquiry method in chemistry laboratories on high school students' conceptions about the nature of science. *Chemistry Education Research and Practice*, *20*(1), 17–29. https://doi.org/10.1039/C8RP00135A

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. Routledge.

Grooms, J., Sampson, V., & Golden, B. (2014). Comparing the effectiveness of verification and inquiry laboratories in supporting undergraduate science students in constructing arguments around socioscientific issues. *International Journal of Science Education*, *36*(9), 1412–1433. http://dx.doi.org/10.1080/09500693.2014.891160

Haro, A. V., Noroozi, O., Biemans, H. J. A., & Mulder, M. (2018). The effects of an online learning environment with worked examples and peer feedback on students' argumentative essay writing and domain-specific knowledge acquisition in the field of biotechnology. *Journal of Biological Education, 53*(4), 390–398, https://doi.org/10.1080/00219266.2018.1472132

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102%2F003465430298487

Koc, C. (2011). The views of prospective class teachers about peer assessment in teaching practice. *Educational Sciences: Theory and Practice*, *11*(4), 1979–1989.

Kolsto, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socioscientific issue. *International Journal of Science Education*, *28*(14), 1689–1716. https://doi.org/10.1080/09500690600560878

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, *77*(4), 339–366. https://doi.org/10.3200/JEXE.77.4.339-366

Lu, J., & Zhang, Z. (2013). Assessing and supporting argumentation with online rubrics. *International Education Studies*, *6*(7), 66–77. http://dx.doi.org/10.5539/ies.v6n7p66

Maloney, J., & Simon, S. (2006). Mapping children's discussion of evidence in science to assess collaboration and argumentation. *Science Education*, *28*(15), 1817–1841. https://doi.org/10.1080/09500690600855419

McDonald, S. (2014). Pre-service primary teachers' written arguments in a socioscientific argumentation task. *Electronic Journal of Science Education*, *18*(7), 1–20.

McLaughlin, P., & Simpson, N. (2004). Peer assessment in first year university: How the students feel. *Studies in Educational Evaluation*, *30*(2), 135–149. https://doi.org/10.1016/j.stueduc.2004.06.003

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook (2nd ed.)*. SAGE Publications.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Noroozi, O., Biemans, H. J. A., & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. *The Internet and Higher Education*, *31,* 20–31. https://doi.org/10.1016/j.iheduc.2016.05.002

Noroozi, O., & Hatami, J. (2018). The effects of online peer feedback and epistemic beliefs on students' argumentation-based learning. *Innovations in Education and Teaching International*, 548–557, https://doi.org/10.1080/14703297.2018.1431143

Noroozi, O., Kirschner, P., Biemans, H. J. A., & Mulder, M. (2017). Promoting argumentation competence: Extending from first- to second- order scaffolding through adaptive fading. *Educational Psychology Review*, 1–24. https://doi.org/10.1007/s10648-017-9400-z

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, *53*(6), 821–846. https://doi.org/10.1002/tea.21316

Sadler, T. D. (2004). Moral and ethical dimensions of socioscientific decision-making as integral components of scientific literacy. *Science Educator*, *13*(1), 39–48.

Sampson, V., Enderle, P., Grooms, J. & Witte, S. (2013). Writing to learn and learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. *Science Education*, *97*(5), 643–670. https://doi.org/10.1002/sce.21069

Sampson, V., Grooms, J., & Walker, J. P. (2011). Argument-Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: an exploratory study. *Science Education*, *95*, 217–257. https://doi.org/10.1002/sce.20421

Sampson, V., & Walker, J. (2012). Argument-Driven Inquiry as a way to help undergraduate students write to learn by learning to write in chemistry. *International Journal of Science Education*, *34*(10), 1443–1485. https://doi.org/10.1080/09500693.2012.667581

Sanmartı´, N. (coord.) (2003). *Aprendre cie`ncies tot aprenent a escriure cie`ncia* [To learn about everything you need to know]. Edicions 62.

Sluijsmans, D., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, *32*, 6–22. https://doi.org/10.1016/j.stueduc.2006.01.005

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessing quality: Criteria for quality assurance in design of (peer) assessment for learning- a review of research studies. *Studies in Educational Evaluation*, *37*, 25–34. https://doi.org/10.1016/j.stueduc.2011.03.004

Topping, K. J., Smith, E. F., Swanson. I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, *25*(2), 149–169. https://doi.org/10.1080/713611428

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.

Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, *47*(8), 952–977. https://doi.org/10.1002/tea.20358

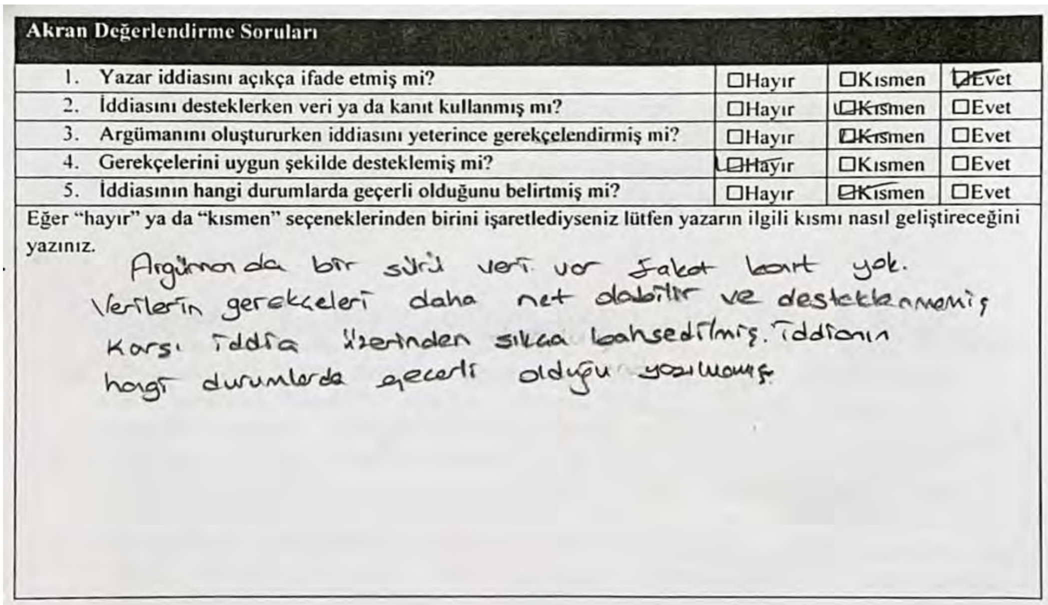
Walker, J. & Sampson, V. (2013). Learning to argue and arguing to learn in science: Argument-Driven Inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching*, *50*(50), 561–596. https://doi.org/10.1002/tea.21082

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Yang, Y. F. (2010). Students' reflection on online self-correction and peer review to improve writing. *Computers and Education*, *55*(3), 1202–1210. https://doi.org/10.1016/j.compedu.2010.05.017

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35–62. https://doi.org/10.1002/tea.10008

Appendix 1. A sample of peer review form in the experimental class.



Akran Değerlendirme Soruları			
1. Yazar iddiasını açıkça ifade etmiş mi?	☐Hayır	☐Kısmen	☒Evet
2. İddiasını desteklerken veri ya da kanıt kullanmış mı?	☐Hayır	☒Kısmen	☐Evet
3. Argümanını oluştururken iddiasını yeterince gerekçelendirmiş mi?	☐Hayır	☒Kısmen	☐Evet
4. Gerekçelerini uygun şekilde desteklemiş mi?	☒Hayır	☐Kısmen	☐Evet
5. İddiasının hangi durumlarda geçerli olduğunu belirtmiş mi?	☐Hayır	☒Kısmen	☐Evet

Eğer "hayır" ya da "kısmen" seçeneklerinden birini işaretlediyseniz lütfen yazarın ilgili kısmı nasıl geliştireceğini yazınız.

Argümanda bir sürü veri var fakat kanıt yok.
Verilerin gerekçeleri daha net olabilir ve desteklenmemiş
Karşı iddia üzerinden sıkça bahsedilmiş. İddianın
hangi durumlarda geçerli olduğu yazılmamış.

Turkish version of a sample peer review form filled out by the one participant in the experimental class.

Peer Review Questions			
1. Did the author explain his/her claim explicitly?	☐ No	☐ Partially	X Yes
2. Did the author use data or evidence to support his/her claim?	☐ No	X Partially	☐ Yes
3. Did the author justify enough the claim when generating his/her argument?	☐ No	X Partially	☐ Yes
4. Did the author support his/her warrants appropriately?	X No	☐ Partially	☐ Yes
5. Did the author specify in which conditions his/her claim was valid?	☐ No	X Partially	☐ Yes

If you selected no or partially, please write how the author will develop the relevant section.

There is a lot of data in the argument but no evidence.
The justifications of the data could be clearer, and it is not supported.
The counterclaim was mentioned frequently.
It is not specified in which cases the claim is valid.

English version of a sample peer review form filled out by the one participant in the experimental class.