MODESTUM
EURASIA Journal of Mathematics, Science and Technology Education, 2022, 18(5), em2104
ISSN:1305-8223 (online)
OPEN ACCESS
Research Paper
https://doi.org/10.29333/ejmste/11966

Evaluation of an elementary teacher education program to promote argument instruction

Tomokazu Yamamoto [1*], Shinichi Kamiyama [2]

[1] Hyogo University of Teacher Education, Hyogo, JAPAN
[2] Osaka University of Health and Sport Sciences, Osaka, JAPAN

Received 14 November 2021 ▪ Accepted 26 March 2022

Abstract
While previous teacher education research on argumentation primarily targeted science teachers, we will provide insights on where to begin teacher education for beginners in situations, where argumentation instruction is not yet common and time constraints impede teacher training. We conducted a short-term program that lasted 160 minutes, with 61 graduate students from a Japanese graduate school of teacher education as participants. It explained arguments, and taught the participants how to construct, score, and plan arguments that they could use in their classes. We conducted the argument construction and evaluation task to examine the presence or absence of components and the correctness of the arguments. The non-parametric quantitative response data were analyzed using SPSS by Wilcoxon signed rank-sum test and Mann-Whitney U test. The results showed that the post-test revealed significant improvements in both tasks. This program may be effective for non-science beginner teachers.

Keywords: argumentation, non-science teachers, short-term teacher education program, argument construction, argument evaluation

INTRODUCTION

This is a research study on teacher education that teaches arguments. Arguments comprise claims, data, warrants, backing, qualifiers, and rebuttals (Toulmin, 1958). As seen in the truism: “science is a social practice and scientific knowledge the product of a community” (Driver et al., 2000, p. 296), it is the persuasion and consensus of evidence-based argumentation that are essential in science lessons. “Not just the theories but even the so-called ‘facts’ of science become argumentative constructions that must be entered into the arena of public debate” (Kuhn, 1993, p. 321). The importance of argumentation has been recognized in science education (e.g., Erduran & Jiménez-Aleixandre, 2007). In PISA 2015, it was noted that one of the areas of scientific literacy is to “interpret data and evidence scientifically–analyze and evaluate data, claims and arguments in a variety of representations, and draw appropriate scientific conclusions” (OECD, 2016, p. 20). There had been a great deal of teaching that introduced argumentation from the elementary school level, mainly in Europe and the United States (e.g., Evagorou et al., 2020; McNeill & Krajcik, 2011; Simon et al., 2011; Zembal-Saul et al., 2012).

To teach argument, it is imperative for teachers themselves to have the opportunity to understand argument and develop their argument construction skills (Iordanou & Constantinou, 2014). Without being able to properly evaluate and provide feedback on students’ arguments, teachers will not be able to develop their abilities. While previous teacher education research on argumentation has mainly focused on science teachers, in Japan, where argumentation instruction is not yet common, and time constraints hinder teacher training, there is a need to understand where to begin teacher education for beginners.

Argument Instruction in Elementary School Science Education

Toulmin’s argument pattern (TAP) has six components, namely: claim, data, warrant, backing, qualifier, and rebuttal (Toulmin, 1958). In recent years, scientific education that applies TAP and introduces argumentation mainly uses data as evidence; warrant

✉ ttyamamo@hyogo-u.ac.jp **(*Correspondence)** ✉ kamiyamashinichi24@gmail.com

Contribution to the literature

- This study provides valuable insights on the effectiveness of a short-term, introductory teacher education program on argument instruction for both in-service teachers who do not specialize in science and elementary pre-service teachers.
- This study is important as it demonstrates program effectiveness in multiple ways by improving argument construction and assessment skills.
- The results of an effective teacher education program can serve as a starting point for elementary school teachers who are not science specialists in implementing argument training in their classrooms in the future.

and backing are integrated as reasoning; and claim, evidence, reasoning, and rebuttal are extracted (e.g., Chin & Osborne, 2010; McNeill & Krajcik, 2011; Sampson et al., 2011). McNeill and Krajcik (2011), who applied TAP to scientific explanations, organized the components so that it was suitable for teaching from the elementary school level onward.

Claim (the answer to a question or a problem), evidence (the scientific data that support the claim), reasoning (the logic behind the choice that articulates why the evidence supports the claim), and rebuttal (a claim about why alternative claims are incorrect, including additional evidence and reasoning to justify that rationale) are the four components of an argument. According to McNeill and Krajcik (2011), evidence is scientific data that supports claim and include information such as observations and measurements from natural settings, as well as results from controlled experiments (p. 23). They can be quantitative or qualitative in nature. Data that has been deemed appropriate and sufficient to justify a claim is used as evidence. They recommend that beginners should ideally compose an argument with three elements, namely: claim, evidence, and reasoning. Our study adopts the McNeill and Krajcik (2011) model, and deals with three beginner components, namely: claim, evidence, and reasoning. This is because they present the components in a way that is easy for children to understand and, similar to this study, introduce the arguments at the elementary school stage.

In addition, the quality of the argument should be assured not only by the presence or absence of these components, but also by the content knowledge (i.e., that it is "correct" in relation to scientific concepts). McNeill and Krajcik (2009), who improved students' argument construction in a middle school chemistry class, have noted that argument construction requires both domain-general support, which is content-independent, and domain-specific support, which is dependent on the content knowledge being addressed. In this study, we will evaluate the quality of the argument from both of these aspects.

Teacher Education for Argumentation Instruction

In order to teach argumentation, a teacher's own ability to construct arguments is essential (Zohar, 2007). However, their ability has been deemed inadequate. Sampson and Blanchard (2012) had science teachers construct an argument that justified one of several explanations for natural events such as the phases of the moon or the color of guppies' bodies. They found that most teachers were unable to support their arguments with the available data. Bilican (2018) explored how pre-service teachers' views of NOS was related with their justifications of SSI and if their justifications changed following an explicit NOS instruction. In the absence of specific instruction on argumentation and SSI, it was reported that although there was an improvement in the NOS view, there was no clear relationship between pre- and post-responses on their sources and variety of justifications. Hence, education programs that enable teachers to construct their arguments are needed.

To teach argumentation, even teachers who do not specialize in science must first recognize the significance and importance of argumentation in science. Erduran et al. (2020) conducted an empirical study on secondary school teachers of science and religious studies in the UK to determine how they perceived the argument and its instruction. They reported that although they all considered the argument to be "important" or "somewhat important," religious studies teachers were more likely than science teachers to incorporate teaching strategies that supported the teaching of the argument. Thus, argument is emphasized in other subjects as well, particularly in those where there is no definitive answer, such as religious studies, rather than in science. As a result, elementary school teachers, especially those who do not specialize in science, should learn more about teaching argumentation. Atabey (2021) has conducted a case study of seven science teachers to investigate the argument types and supporting reasons that they presented in the context of COVID-19, the current SSI. They found that science teachers supported their decisions regarding SSI in the context of COVID-19. They also found that science teachers use personal experiences, values, social conditions, and health factors to support their decisions about SSI in the context of COVID-19 from multiple subject perspectives. This study suggests that when training teachers in argument construction, a cross-curricular approach that is not limited to science is needed to help them make decisions on real-life issues such as SSI.

To improve teachers' ability to construct arguments, for example, in a study of in-service teachers in elementary schools, Zembal-Saul (2009), through a project called teaching elementary school science as an argument (TESSA) had teachers watch science lesson videos that examined the relationship between air and air pressure. She analyzed their responses to a series of questions on the video and reported that she could make in-service teachers aware of the importance of letting students construct scientific explanations from evidence.

In addition, research has highlighted the effectiveness of teacher education programs that allow teachers to construct their arguments. For example, Iordanou and Constantinou (2014) practiced the activity of using evidence to facilitate arguments between pairs with opposing views among Cypriot pre-service teachers who were enrolled in a mandatory science education course, and reported that they were able to use evidence to construct arguments and that the teachers themselves were aware of this at a meta level. Kaya (2013) asked pre-service teachers to construct an argument in the chemistry domain and compared the results between lecture-based control and experimental groups where argumentation activities were incorporated in the class. The results showed that the latter group had a higher level of argumentation skills and greater understanding of the content, and emphasizing the importance of teacher education in shaping argumentation skills among teachers.

Furthermore, teachers who introduce argumentation into their classes must have the ability to evaluate learners' argumentation appropriately. Osborne et al. (2004) incorporated activities in their teacher education program where the teachers constructed and evaluated arguments. Teachers referred to samples of students' arguments and evaluated each component by checking for the claim and whether the data supported the claim. In their instructional guide for teachers, McNeill and Krajcik (2011) guided teachers in creating evaluation tasks based on science curriculum standards and levels of complexity of arguments. They also provided an example of a rubric for evaluating students' arguments in a unit on food chains and chemical reactions for middle school students, in terms of the presence of claims, evidence, and reasoning, as well as their appropriateness and sufficiency.

In Japan, although the importance of argumentation has been pointed out in response to the results of PISA 2015, the need for and understanding argument instruction has not been widespread. Therefore, it is necessary to recognize that argument instruction is necessary first. In this process, it is important to acquire the ability to construct and evaluate arguments as described above. These are the skills that teachers who teach arguments should acquire first as beginners. This is due to the risk of referring to an existing discussion or other instruction as an argument if the entire structure of the argument is not understood (Katsh-Singer et al., 2016; McNeill et al., 2016). However, previous teacher education research has primarily focused on science teachers, and the programs have been relatively long-term. As a result, there are some challenges in introducing them because they are designed for novice teachers who do not specialize in science. In addition, most programs focus on the ability to construct arguments by focusing on the components. There are also no short programs that include activities to evaluate learners' arguments from the same viewpoint or envision the arguments that they would like to construct in actual classroom situations. There is a need for an introductory short-term program for elementary school teachers who are not science specialists and are busy teaching a variety of subjects to become interested in teaching argumentation in their science classes. The aforementioned studies either targeted teachers who specialized in science, or examined the effects of teacher education only on one aspect, such as argument comprehension, construction, or evaluation. The program effectiveness should be evaluated from the teachers' ability to both construct and evaluate arguments, targeting those who do not specialize in science.

Yamamoto and Kamiyama (2018) developed an introductory program for teachers who do not specialize in science and elementary school teachers in Japan to introduce argumentation in their classrooms. This 90-minute program includes an explanation of argumentation and the methods to construct and evaluate arguments. It improved the participants' ability to construct and evaluate arguments by making them aware of the significance and various components of arguments. However, this program was mainly lecture-based, and it was difficult to visualize concretely how to introduce the argument into their own classes.

The Purpose of the Study

This study examines how providing a beginners' program to elementary school teachers and other non-science specialists help them develop the most basic argument construction and evaluation skills while focusing on the components. Previous teacher education research on argumentation mainly targeted science teachers; however, in Japan, argumentation instruction is not yet common and time constraints hinder teacher training. Hence, we will provide insights on where to begin teacher education for beginners.

This study implemented a short-term teacher education program for non-science teachers to introduce argumentation for the first time in their classes and to

Table 1. The details of the 35 in-service teachers

Category		Number of people
Sex	Male	26
	Female	9
School	Elementary school	18
	Junior high school (Physical education, 2; National language, 1; Social studies, 1; Mathematics, 1; English, 1; Art, 1; Technology, 1)	8
	Senior high school (National language, 2; English, 2; Geography and history, 1; Commerce, 1; Fisheries, 1)	7
	Special needs school	2
Age	20s	1
	30s	16
	40s	17
	50s	1
Teaching experience	Under 10 years	5
	10 to 20 years	19
	20 to 30 years	11

examine the program's effects as improvement in the teachers' ability to construct and evaluate arguments. This study sought to answer two research questions:

1. Did the introductory short-term program to introduce argumentation into elementary science classes *improve the argument construction and evaluation skills* of in-service teachers who do not specialize in science and elementary pre-service teachers?
2. Are there *differences in the program's effects between both groups*?

METHODOLOGY

This study is a small-scale investigation of a practical study on teacher education. This method was used because this study is a preliminary short program that aims to provide practical findings for its future improvement. We will also find directions for future research by clarifying the teachers' ability to construct and evaluate arguments as an effect of the developed teacher education program.

Subjects

This study was conducted during the period of the Graduate School of Teaching' required course called "Seminar on Lesson Instructional Planning and Research of Teaching Materials." Here, both graduate students who are in-service teachers at elementary, middle, and high schools, as well as graduate students who wish to become elementary school teachers take the same course. The instructor was the author, who devoted 160 minutes of his time (not including the questionnaire survey) to this program.

A total of 61 graduate students from a Japanese graduate school for teacher education participated in this study. All of them had no experience in teaching argumentation. Of them, 26 were pre-service teachers who wanted to become elementary school teachers (16 men, 10 women). They all attended the program in April-May 2017. Table 1 provides the details of the remaining 35 in-service teachers. 35 (26 men, nine women; 18 elementary school teachers; eight junior high school teachers, seven senior high school teachers, and two special needs teachers were in-service teachers who did not specialize in science, and one teacher was in his 20s, 16 were in their 30s, 17 were in their 40s, and one was in his 50s. There were five teachers with less than 10 years of teaching experience, 19 teachers with 10 to 20 years of teaching experience, and 11 teachers with 20 to 30 years of teaching experience. Of the 35 in-service teachers, 12 attended the program in April-May 2017 and the remaining 23 attended the same program in April-May 2018.

Program

Table 2 presents the outline of the program. Activity 1 was a lecture on the definition and significance of argumentation (20 minutes). We explained that language activities are emphasized in "the course of study" in Japan (MEXT, 2010) and that scientific knowledge is provisional, and introduced Toulmin's (1958) argument as a method of scientific explanation to build consensus in science.

We gave concrete examples of how evidence-based argumentation is emphasized in PISA and domestic academic achievement tests, so that the teachers can understand why argumentation is necessary. Activity 2 comprised a lecture and exercises on children's actual arguments in writing (20 minutes). Here, we presented three examples of arguments written by sixth-grade students in response to a question on the weather and temperature (Appendix A). They were asked to judge whether the claim, evidence, and reasoning were present in each argument. Then, they were asked to judge whether the arguments were correct or not, based on the rubric.

Table 2. The outline of the program

Activity1 (20 minutes)
Lecture on the definition and significance of argument.
Activity2 (20 minutes)
Lecture and exercises on children's actual writing arguments.
Activity3 (40 minutes)
Exercises to gain an experiential understanding of teaching and evaluating arguments.
Activity4 (10 minutes)
An overview of actual arugment instruction in elementary school classes.
Activity5 (30 minutes)
Develop a plan to implement the argument into own class.
Activity6 (40 minutes)
Exchange and review of teaching plans.

Table 3. The rubric for correctness of contents

Score	Claim	Evidence	Reasoning
0	There is no description of claim.	There is no description of evidence.	There is no description of reasoning.
	The wrong claim is described.	The wrong evidence is described.	The wrong reasoning is described.
1	One of the following two is described.	One of the following two is described.	One of the following two is described.
	1. A applies to October 10.	1. In A, the daily temperature change is 6°C. Or, the daily temperature is between 18°C and 24°C.	1. Because the temperature on a cloudy day does not change much throughout the day.
	2. B applies to October 20.	2. In B, the daily temperature change is 11°C. Or the daily temperature is between 13°C. and 24°C.	2. Because the temperature on a sunny day is low in the morning and evening and high around noon, and changes greatly throughout the day.
2	Both of above contents are described.	Both of above contents are described.	Both of above contents are described.

Table 4. The child's argument and correctness judgement

Example	Children's descriptions	Presence/absence of description	Correctness of contents
1	The temperature in A is between 18°C and 24°C throughout the day, while the temperature in B is between 13°C and 24°C. So A is the temperature on October 10, and B is the temperature on October 20. On a sunny day, the temperature changes significantly throughout the day, and on a cloudy day, there is no significant change throughout the day. So A has no significant change in temperature, while B has a significant change.	Claim=2, Evidence=2, Reasoning=2	Claim=2, Evidence=2, Reasoning=2
2	I think October 10 is temperature A, and October 20 is temperature B. The reason is that if you make a line graph, on a sunny day you will see a mountain shaped graph with a large change in temperature, but on a cloudy day you will see a graph with little change in temperature.	Claim=2, Evidence=0, Reasoning=2	Claim=2, Evidence=0, Reasoning=2
3	I believe that B is cloudy and A is sunny because B is generally cooler and A is generally warmer.	Claim=0, Evidence=2, Reasoning=0	Claim=0, Evidence=0, Reasoning=0

Table 3 presents the rubric for the correctness of the contents. Table 4 shows the child's argument and correctness judgment. Example 1 is where all the components are present and the contents are correct. In example 2, there is no mention of the data on temperature change in A and B, and the case lacks evidence. In example 3, the question asks the respondent to answer whether it is 10 or 20 days, but the respondent answers sunny or cloudy, and the claim does not directly address the question (there is no part that addresses the claim). It mentions temperature changes in A and B as evidence, but does not mention numerical data and is incomplete. The temperature change because of weather is not mentioned, and there is no reasoning in this case. We conducted an activity to share these judgment methods and results.

In activity 3, titled "exercises to gain an experiential understanding of teaching and evaluating arguments" (40 minutes), students were asked to describe an argument comprising a claim, evidence, and reasoning. We also introduced variations in the complexity of an argument as devised by McNeill and Krajcik (2011). The science lesson was conducted in the fifth grade of elementary school, and the theme was "motion of a pendulum." The graduate students conducted an experiment to answer the question, "can you make the period of a pendulum one second by changing the

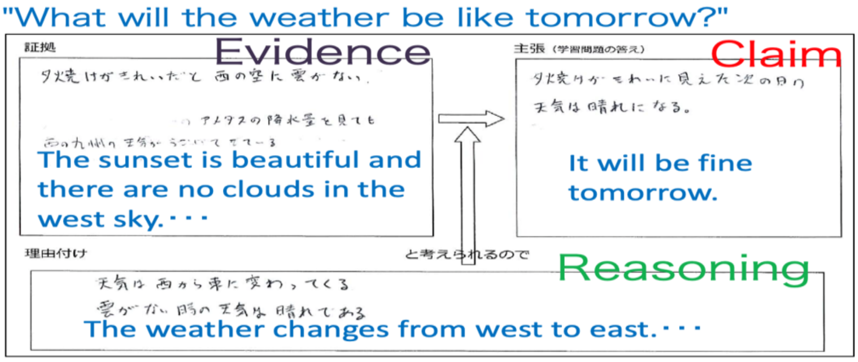


Figure 1. An argument created by a graduate student

weight?" When the results of the experiment were presented, the teachers argued, "it is not possible to make the period of the pendulum one second by changing the weight of the weight," and gave evidence, saying that "the period of the pendulum with both 10 g and 20 g weights was the same, with an average of 1.4 seconds." They also offered the following reasoning: "the weight of the weight is not related to the period."

In activity 4, titled "an overview of the actual argument instruction in elementary school classes" (10 minutes), we provided an overview of the actual class and explained the teaching strategy for introducing argumentation into the class. We showed that the activity improved the children's ability to construct arguments and their understanding of the learning content.

The last activity called on the teachers to devise an argument introduced in their classrooms. Activity 5 was titled "develop a plan to implement the argument in your class" (30 minutes) and activity 6 was titled "exchange and review of teaching plans" (40 minutes). If they had to introduce an argument in their class, they were asked to envision the unit and the kind of questions, claims, evidence, and reasoning they would use, and to create a one-hour lesson plan based on this. Then, they were asked to exchange their lesson plans and evaluate each other's claims, evidence, and reasoning, and whether or not they were appropriate. Figure 1 shows an argument created by one graduate student. This worksheet was created by referring to the argument diagram in Chin and Osborne (2010), and is presented with associated boxes for filling in claims, evidence, and reasoning. In response to the question "what will the weather be like tomorrow?", the graduate student constructed the following argument: "it will be fine tomorrow (claim), the sunset is beautiful and there are no clouds in the western sky (evidence), the weather changes from west to east (reasoning)."

Data Collection

Before and after the program, two surveys were administered to the learners. The first was an argument construction task (Appendix A), where the question was taken from Sakamoto et ai. (2012). The second was an argument evaluation task (Appendix B), which examines whether the teacher can appropriately evaluate an argument constructed by the learner based on the existence and correctness of the claim, evidence, and reasoning components. Each task took about 10 minutes to complete. The subjects are 35 in-service teachers who did not specialize in science and 26 pre-service teachers who wanted to become elementary school teachers, all of whom participated in the program. We explained to the participants before the program that the tasks would be conducted as part of the class and that they would be used in a manner that would not identify any of them. We obtained a full written consent from all participants.

Argument Construction Task

Figure 2 and Figure 3 present the question and desirable answer on argument construction, respectively. Teachers were asked to construct an argument, a question wherein a circuit in a hidden area was to be selected from between two choices. Teachers were assigned a score out of two points in three components (for a total of six points), based on a rubric (Appendix B) evaluating "whether an answer was given (presence/absence of component)" and "whether the answer was scientifically correct (correctness of content)."

The three components were as follows: their "claim" as an answer to the question, their "evidence" mentioning the brightness of miniature light bulbs in each circuit (the fact of the matter), and their "reasoning" mentioning the size of each current (voltage/power) connected in parallel. We counted the number of people by score for each component. Two independent judges evaluated the answers with a concordance rate of 97.8%.

Argument Evaluation Task

Figure 4 presents the argument evaluation question. Teachers were asked to evaluate answers from children X, Y, and Z (cases X, Y, and Z), who were each presented with a U-shaped magnet of unknown polarity and asked to determine the polarity by bringing the north pole of a bar magnet in contact with it.

The circuit of ① and ② has the part of the cell hidden.
In the circuit of ①, the brightness of the light bulb was brighter than that of one cell. In the circuit of ②, the brightness of the light bulb was the same as that of one cell.
Which of the following A and B applies to how to connect the cells of each of the circuits 1 and 2? And why did you think so?
Explain these things scientifically. Write an explanation.

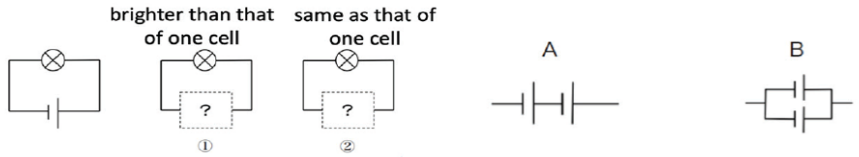
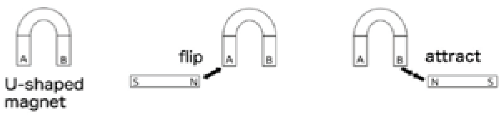


Figure 2. Argument construction question

Claim: ① is A, ② is B.
Evidence: In the circuit of ①, the brightness of the light bulb was brighter than that of one cell. In the circuit of ②, the brightness of the light bulb was the same as that of one cell.
Reasoning: If you increase the number of cells to two in series connection and turn on the light bulb, the current (or any of voltage, power and electricity) will increase compared to when connected to one cell.
Even if you increase the number of cells to two and connect the light bulb with parallel connection, the current (or any of voltage, power, and electric quantity) does not change compared to the time when one cell is connected.

Figure 3. Desirable answer in the argument construction question

What are the poles of A and B of the U-shaped magnet? And why did you think so? Explain these things scientifically.



If you can judge that the children's descriptions X, Y and Z are sufficient for scientific explanation, enter ○ in the "judgment" column. If you judge that it is not enough, enter △. And briefly state in the "reason" column why you have judged that.

【Child X】
judgment [ ]
reason [ ]
A is N pole and B is S pole.
When the N pole was brought close to A, the magnets was flipped, and when the N pole was brought close to B, the magnets attracted each other.

【Child Y】
judgment [ ]
reason [ ]
When the N pole was brought close to A, the magnets was flipped, and when the N pole was brought close to B, the magnets attracted each other.
The magnets have the same poles are flipped and the other poles attract each other.

【Child Z】
judgment [ ]
reason [ ]
A is N pole and B is S pole.
The magnets have the same poles are flipped and the other poles attract each other.

Figure 4. Argument evaluation question

Table 5. Score distribution for the argument construction (in-service teachers)

		Pre-test			Post-test			
	Score	2	1	0	2	1	0	
Presence/ absence of component	Claim	32	1	2	35	0	0	
	Evidence	13	0	22	**26**	**3**	**6**	**
	Reasoning	27	3	5	32	1	2	
Correctness of contents	Claim	30	1	4	33	0	2	
	Evidence	14	0	21	**26**	**2**	**7**	**
	Reasoning	19	3	13	20	1	14	

Note. Components showing significant improvement in the score distribution between pre- & post-test are shown in **bold**

Table 6. Score distribution for the argument construction (pre-service teachers)

		Pre-test			Post-test			
	Score	2	1	0	2	1	0	
Presence/ absence of component	Claim	25	0	1	24	0	2	
	Evidence	6	0	20	**12**	**2**	**12**	*
	Reasoning	23	1	2	22	3	1	
Correctness of contents	Claim	21	0	5	23	1	2	
	Evidence	6	0	20	**12**	**2**	**12**	*
	Reasoning	15	1	10	13	1	12	

Note. Components showing significant improvement in the score distribution between pre- & post-test are shown in **bold**

In judging these answers, the teachers were asked to focus on the students' "claim," that is, the answers they gave for each pole; "evidence" mentioning how the bar magnet reacted in the way that it did (the facts of the matter); and "reasoning," indicating that magnets of the same polarity are repulsed whereas magnets of opposite polarity are attracted. Each case was an incomplete argument, lacking either reasoning, claim, or evidence.

For each argument, in the "judgment" section, one point was given to point out incorrectness. In the "reasons" section, one point was given if the lack of a component was mentioned. In all other cases, the score was 0 (Appendix C). We counted the number of people by viewpoints score for each case. Two independent judges evaluated these answers with a concordance rate of 99.6%.

Rubrics and Analysis Tools

In order to examine the presence or absence of components and the correctness of the arguments described by the students, we used the rubric of argument construction skills by Sakamoto et al. (2012), and the rubric of evaluation skills developed based on it, to score the students' arguments. The rubric for the argument construction task (Sakamoto et al., 2012) was developed by three university faculty members specializing in science education and two university faculty members specializing in psychology. It has mainly been used in analyses of in-service teachers and elementary school students (Sakamoto et al., 2012), all with agreement rates of 90% or higher. Meanwhile, the rubric for the argument evaluation task was developed by two university faculty members specializing in science education and has been used in analyses of in-service teachers (Yamamoto & Kamiyama, 2017), both with 90% or higher agreement. The quantitative response data was analyzed using IBM SPSS statistics v 27, where the Wilcoxon signed rank-sum test and Mann-Whitney U test were selected because the data set is non-parametric.

RESULTS

Argument Construction Task

Table 5 and Table 6 present the distribution of scores for the argument construction task. The Wilcoxon signed rank-sum test results showed a significant improvement in the distribution of scores between the pre- and post-tests. In-pre-service teachers had mostly perfect scores on the pre-post-tests for "claim," both for the presence/absence of the component and the correctness of the content. Both in- and pre-service teachers showed significant improvements in the presence/absence of the component and the correctness of the content (in-service teachers' presence/absence of component, $Z=3.266$, $p<.01$; evidence, $Z=2.887$, $p<.01$; pre-service teachers' presence/absence of component, $Z=2.269$, $p<.05$; evidence, $Z=2.269$, $p<.05$) for "evidence." In-pre-service teachers received mostly perfect scores on both the pre-post-test in the presence/absence of the component for "reasoning," whereas only about half received perfect scores in terms of the correctness of the content.

Argument Evaluation Task

Table 7 and Table 8 present the distribution of scores for the argument evaluation task. The results of the McNemar test showed that for both in- and pre-service teachers, cases X and Y, where "reasoning" and "claim" were lacking, respectively, both "judgment" and "reasoning" had mostly perfect scores in the post-test. In case Z, where evidence was lacking, the number of perfect scores increased significantly from the pre- to the

Table 7. Score distribution for the argument evaluation (in-service teachers)

		Pre-test		Post-test		
	Score	1	0	1	0	
Judgment	Case X	32	3	35	0	
	Case Y	32	3	35	0	
	Case Z	16	19	**28**	**7**	**
Reason	Case X	28	7	34	1	
	Case Y	30	5	34	1	
	Case Z	11	24	**26**	**9**	**

Note. Components showing significant improvement in the score distribution between pre- & post-test are shown in **bold**

Table 8. Score distribution for the argument evaluation (pre-service teachers)

		Pre-test		Post-test		
	Score	1	0	1	0	
Judgment	Case X	23	3	24	2	
	Case Y	23	3	26	0	
	Case Z	8	18	**24**	**2**	***
Reason	Case X	22	4	23	3	
	Case Y	21	5	24	2	
	Case Z	7	19	**21**	**5**	***

Note. Components showing significant improvement in the score distribution between pre- & post-test are shown in **bold**

post-test stages (in-service teachers' judgment and reason, $p<.01$; pre-service teachers' judgment and reason, $p<.001$).

DISCUSSION

Ability to Construct and Evaluate Arguments

Did the short-term introductory program to introduce arguments into elementary science classes improve the argument construction and evaluation skills of in-service teachers who did not specialize in science and elementary pre-service teachers? The program resulted in generally good scores in the post-test for the argument construction and evaluation tasks, and the teachers were able to write, check, and evaluate their arguments with a strong theoretical awareness. This can be attributed to the program's activity, where the teachers themselves constructed arguments and scored the learners' arguments. The teachers could construct and evaluate arguments appropriately, even when the content changed. Lytzerinou and Iordanou (2020, p. 631) noted:

> A connection between teachers' ability to evaluate arguments and their own skills of constructing arguments, has important educational implications, suggesting that teachers' own argument skills play a key role in their ability to evaluate arguments which is closely connected with teachers' ability to support the development of students' argument skills.

The program had a certain effect on constructing and evaluating arguments. However, some graduate students encountered problems in providing evidence and justifying their arguments accurately. They tended to consider the bulb brightness in each circuit (a fact) as self-evident and omitted it altogether. This kind of incomplete argument is also seen among elementary school students. McNeill and Krajcik (2011) reported a case where a fifth-grade student determined the population of hawks when all seeds were removed from an ecosystem, but the evidence to support his claim was insufficient. When Zembal-Saul et al. (2012) asked fourth-grade students about the effect of washers on the speed of a car, the students could not provide adequate evidence or reasoning. When graduate students construct their arguments, they need to use a lot of self- and peer-assessment methods that check the presence or absence of components and their correctness.

There were some cases where the students failed to mention the amount of current (voltage and power) in the series and parallel connection of dry cell batteries, which caused a difference in the brightness of the light bulbs. This is probably because the graduate students were not specialized in science and did not think of the interpretation of these scientific terms. They may have taken particular phenomena (facts) such as "connecting dry cell batteries in series makes a light bulb brighter" as their reasoning. In addressing this state of confusion between facts and reasoning through scientific principles (interpretation), Berland and Reiser (2009) pointed out that it is difficult for middle school students to distinguish between the components of an argument. The confusion between evidence and reasoning mentioned in the TIMSS 2003 overlapped with the problem of learners' difficulties in distinguishing between facts and reasons. This problem may be influenced by the (lack of) content knowledge of graduate students who do not specialize in science. Sadler and Fowler (2006) evaluated the arguments of high school students and junior college students majoring in non-science subjects and in the subjects of gene therapy and cloning, and reported that the junior college students majoring in science frequently referred to the content knowledge of science in their justifications and constructed excellent arguments. In contrast, Faize et al. (2018) noted that developing arguments with graduate students who have no prior knowledge or hold contradictory beliefs can lead to problems accepting their arguments and creating confrontational situations in class. For anyone to compose an effective argument to persuade others, it is important to understand the correct scientific knowledge that can be used as a rationale.

The Difference Between In- and Pre-service Teachers

Are there differences in the program's effects between both groups? In- and pre-service teachers are similar in their ability to construct and evaluate arguments. This may be related to the fact that both share non-specialization in science and have no prior experience teaching argumentation. This shows that regardless of their career status, a special teacher education program is essential to teach argumentation. In a study of teacher education specializing in science, Aydeniz and Ozdilek (2015) conducted an 11-week program for 40 in-service science teachers in Turkey that included activities for constructing, evaluating, and critiquing arguments, teaching three argumentation lessons, peer-observations of lessons, and opportunities for reflection on teaching skills. From their analysis results, they inferred that the mastery over constructing scientific arguments in class, and evaluating the quality of and critiquing other's arguments and lessons contributed improving of the participants' self-efficacy. Such a program can be effective for in-service teachers who do not specialize in science and pre-service teachers who wish to become elementary school teachers.

IMPLICATION

Previous teacher education research on argumentation mainly targeted science teachers and often involved programs lasting several weeks (e.g., Aydeniz & Ozdilek, 2015; Iordanou & Constantinou, 2014; Kaya, 2013; Osborne et al., 2004). In a place like Japan, where argument instruction is not yet common and time constraints hinder teacher training, there is a lack of knowledge on where to start teacher education for beginners. In this study, we clarified the effectiveness of a short-term program as an initial step for teachers to recognize the significance of argumentation, introduce argumentation into their own practice, and verify their results. The program was able to improve teachers' ability to construct and evaluate arguments to a certain extent, even within a constrained time of 160 minutes. The same effect was observed for both pre-service and in-service teachers with extensive teaching experience.

This achievement, though limited, may have encouraged the meta-level awareness of the argument construction reported by Iordanou and Constantinou (2014). It also has the potential to lead to the content understanding of science concepts achieved by Kaya (2013) and to the teachers' belief that they should learn to construct and evaluate arguments, as argued by Aydeniz and Ozdilek (2015). Improving the ability to construct and evaluate arguments by focusing on the components of this study is an important progression at the early stage. In teaching argumentation, it is required to recognize the importance of the argumentation in the context of logically persuading others to agree based on evidence. In order to do this, it is necessary to focus on the components of an argument, check for their existence and correctness, and construct an argument, while at the same time evaluate the learners' arguments from a consistent perspective. Based on these activities, it is possible for teachers who do not specialize in science to have a starting point for implementing argumentation training in their classrooms in the future.

In addition, non-science teachers (whether experienced in-service or pre-service) had the same problems in argument construction and evaluation, which was improved by the program. The ability of teachers to construct and evaluate arguments does not come naturally through teaching experience, but requires deliberate training. This is consistent with the views of many researchers, including Bilican (2018) and Zoher (2007).

Limitations of the Study

Because this study was only a small-scale survey, conclusions about the program's effects were limited. Data from a larger survey of non-science teachers are needed. It should also be clarified at that point what differences exist in the argument by subject. This study also fails to answer the question on how the program has transformed the actual teaching of classroom teachers, especially regarding how it has affected their self-efficacy in teaching argumentation. It is necessary to confirm the effect on the improvement of teachers' teaching ability. There is a need for a case study that targets teachers who became interested in argument instruction through this program, and the subsequent changes in their teaching and the impact on their self-efficacy.

Future Research

Based on these results, future tasks include examining the characteristics of argumentation in a variety of subjects and implementing the program with a larger sample to clarify the improvement of teachers' argumentation construction and evaluation skills. Furthermore, with regard to self-efficacy, Lytzerinou and Iordanou (2020) had social and physical science teachers construct an argument on history and socio-scientific topics, and reported that the teachers demonstrated the same level of self-efficacy in teaching argumentation, whether in their own fields of expertise or in other fields. Even for elementary school teachers who do not specialize in science, a teacher education program on argumentation can improve their self-efficacy in science classes, independent of the domain. Aydeniz and Ozdilek (2015, p. 1271) discussed in-service teachers increasing their self-efficacy to teach science through argumentation. They suggested that teachers should understand argumentation both as a scientific practice and as a pedagogical tool, and that they should learn to construct and evaluate arguments. They also indicate the importance of having these experiences over a long period and reflecting on the experiences with a group of teachers.

Based on these findings, future program development after the introductory period should include construction, evaluation, and lesson planning for argumentation, and ensure the actual introduction and reflection of argumentation in the classroom. In the case of elementary school teachers, who do not specialize in science, it may be better to introduce argumentation in their specialized subjects first, rather than in science. This experience will give them the confidence to introduce argumentation in elementary science classes. The challenge is to reflect on the teaching of argumentation through concrete examples in practice, based on the abilities that were improved in this study, and strengthen confidence.

The teachers' own learning approaches and beliefs also have a significant impact on their ability to construct and evaluate argument. Teachers with deeper learning approach, willingness to learn, intense and critical interaction with content, and ability to link previous knowledge with new learnings, associate concepts with daily experiences, and establish a relationship between events and outcomes, have the ability to build better arguments (Aydogan et al., 2017). It has also been noted that in introducing argument instruction, teachers' beliefs about students' science learning and ability, their role as a science teacher, and contextual factors such as district and standards, had significant influences on how they interpreted and adapted an instructional model (Sengul et al., 2021). Given these considerations, the next step is to expand on the short-term program introduced in this study to create a long-term intensive program that includes a deeper learning approach and promotes a stronger belief in the importance of argument.

Author contributions: All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: This research was supported by JSPS KAKENHI Grant Number 18K02936.

Declaration of interest: No conflict of interest is declared by authors.

REFERENCES

Atabey, N. (2021). Science teachers' argument types and supporting reasons on socioscientific issues: COVID-19 pandemic. *International Journal of Psychology and Educational Studies*, *8*(2), 214-231. https://doi.org/10.52380/ijpes.2021.8.2.500

Aydeniz, M., & Ozdilek, Z. (2015). Assessing and enhancing pre-service science teachers' self-efficacy to teach science through argumentation: Challenges and possible solutions. *International Journal of Science and Mathematics Education*, *13*(3), 1-19. https://doi.org/10.1007/s10763-015-9649-y

Aydogan, N., Polat, H., Cankaya, O., & Emre, F. B. (2017). Effect of learning approaches of pre-service science teachers on the argument create skills. *SHS Web of Conferences*, *37*, 01044. https://doi.org/10.1051/shsconf/20173701044

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26-55. https://doi.org/10.1002/sce.20286

Bilican, K. (2018). Analysis of pre-service science teachers' understanding of nature of science and proposed arguments on socio-scientific issues. *International Journal of Research in Education and Science*, *4*(2), 420-435. https://doi.org/10.21890/ijres.410632

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *The Journal of the Leaning Sciences*, *19*(2), 230-284. https://doi.org/10.1080/10508400903530036

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287-312. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Erduran, S., & Jiménez-Aleixandre, M. P. (Eds.). (2007). *Argumentation in science education: Perspectives from classroom-based research*. Springer. https://doi.org/10.1007/978-1-4020-6670-2

Erduran, S., Guilfoyle, L., & Park, W. (2020). Science and religious education teachers' views of argumentation and its teaching. *Research in Science Education*, *52*(4), 1-19. https://doi.org/10.1007/s11165-020-09966-2

Evagorou, M., Nicolaou, C., & Lymbouridou, C. (2020). Modelling and argumentation with elementary school students. *Canadian Journal of Science, Mathematics and Technology Education*, *20*, 58-73. https://doi.org/10.1007/s42330-020-00076-9

Faize, F. A., Husain, W., & Nisar, F. (2018). A critical review of scientific argumentation in science education. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 475-483. https://doi.org/10.12973/ejmste/80353

Iordanou, K., & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction*, *34*, 42-57. https://doi.org/10.1016/j.learninstruc.2014.07.004

Katsh-Singer, R., McNeill, K. L., & Loper, S. (2016). Scientific argumentation for all? Comparing teacher beliefs about argumentation in high, mid, and low socioeconomic status schools. *Science Education*, *100*(3), 410-436. https://doi.org/10.1002/sce.21214

Kaya, E. (2013). Argumentation practices in classroom: Pre-service teachers' conceptual understanding of chemical equilibrium. *International Journal of Science Education*, *35*(7), 1139-1158. https://doi.org/10.1080/09500693.2013.770935

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319-337. https://doi.org/10.1002/sce.3730770306

Lytzerinou, E., & Iordanou, C. (2020). Teachers' ability to construct arguments, but not their perceived self-efficacy of teaching, predicts their ability to evaluate arguments. *International Journal of Science Education*, *42*(4), 617-634. https://doi.org/10.1080/09500693.2020.1722864

McNeill, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Science, 18*(3), 416-460. https://doi.org/10.1080/10508400903013488

McNeill, K. L., & Krajcik, J. (2011). *Supporting grade 5-8 student in constructing explanation in science.* Pearson.

McNeill, K. L., Katsh-Singer, R., González-Howard, M., & Loper, S. (2016). Factors impacting teachers' argumentation instruction in their science classrooms. *International Journal of Science Education, 38*(12), 2026-2046. https://doi.org/10.1080/09500693.2016.1221547

MEXT. (2010). Course of study for elementary schools. *Ministry of Education, Culture, Sports, Science and Technology*. https://www.mext.go.jp/content/20201008-mxt_kyoiku02-000005241_1.pdf

OECD. (2016), *PISA 2015 Assessment and analytical framework: Science, reading, mathematic and financial literacy*. OECD Publishing. https://doi.org/10.1787/9789264255425-en

Osborne, J., Erduran, S., & Simon, S. (2004). *Ideas, evidence and argument in science.* King`s College, London.

Sadler, T. D., & Fowler, S. R. (2006). A threshold model of content knowledge transfer for socioscientific argumentation. *Science Education, 90*(6), 986-1004. https://doi.org/10.1002/sce.20165

Sakamoto, M., Yamamoto, T., Yamaguchi, E., Nishigaki, J., Muratsu, K., & Inagaki, S. (2012). Argument skill acquisition among grade 5 and grade 6 elementary school students. *Journal of Science Education in Japan, 36*(3), 252-261. https://doi.org/10.14935/jssej.36.252

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching, 49*(9), 1122-1148. https://doi.org/10.1002/tea.21037

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*(2), 217-257. https://doi.org/10.1002/sce.20421

Sengul, O., Enderle, P. J., & Schwartz, R. (2021). Examining science teachers' enactment of argument-driven inquiry (ADI) instructional model. *International Journal of Science Education, 43*(8), 1273-1291. https://doi.org/10.1080/09500693.2021.1908641

Simon, S., Johnson, S., Cavell, S., & Parsons, T. (2011). Promoting argumentation in primary science contexts: An analysis of students' interactions in formal and informal learning environments. *Journal of Computer Assisted Learning, 28*(5), 440-453. https://doi.org/10.1111/j.1365-2729.2011.00451.x

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

Yamamoto, T., & Kamiyama, S. (2017). Developing an elementary school teacher training program for improving argument construction and evaluation skills. *Journal of Research in Science Education, 57*(4), 387-401. https://doi.org/10.11639/sjst.sp16009

Yamamoto, T., & Kamiyama, S. (2018). Teacher beliefs about argumentation in Japanese in-service teachers. In O. Finlayson, E. McLoughlin, S. Erduran, & P. Childs (Eds.), *Electronic Proceedings of the ESERA 2017 Conference. Research, Practice and Collaboration in Science Education, Part 7* (co-ed. M. Andrée & J. Viiri), (pp. 952-959.). Dublin City University. https://www.researchgate.net/profile/Anne-Freitas-2/publication/330938030_ALCESTE_SOFTWARE_USAGE_IN_THE_IDENTIFICATION_OF_SPEECHES_FOUND_IN_WRITTEN_TEXTS_ABOUT_THE_DIGESTIVE_SYSTEM/links/5c62f6a9a6fdccb608be29df/ALCESTE-SOFTWARE-USAGE-IN-THE-IDENTIFICATION-OF-SPEECHES-FOUND-IN-WRITTEN-TEXTS-ABOUT-THE-DIGESTIVE-SYSTEM.pdf#page=47

Zembal-Saul, C. L. (2009). Learning to teach elementary school science as argument. *Science Education, 93*(4), 687-719. https://doi.org/10.1002/sce.20325

Zembal-Saul, C. L., McNeill, K. L., & Hershberger, K. (2012). *What's your evidence? Engaging K-5 children in constructing explanations in science.* Pearson.

Zohar, A. (2007). Science teacher education and professional development in argumentation. In S. Erduran, & M. P. Jiménez-Alexandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 245-268). Springer. https://doi.org/10.1007/978-1-4020-6670-2_12

APPENDIX A

The weather on the playground was "cloudy" on October 10 and "sunny" on October 20.

Which of the two temperatures in Table A applies to the daily temperature change on October 10 and October 20, A or B? And why do you think so?

Please explain these things in scientific terms.

Table A. Daily temperature change in the playground

	9AM	10AM	11AM	Noon	1PM	2PM	3PM	4PM
Temperature A (°C)	18	18	20	21	23	24	23	23
Temperature B (°C)	13	16	18	21	23	24	22	20

Please write your explanation below.

APPENDIX B

Table B. Correctness of contents rubric for the argument construction task

Score	Claim	Evidence	Reasoning
0	There is no description of the claim.	There is no description of the evidence.	There is no description of the reasoning.
	The wrong claim is described.	The wrong evidence is described.	The wrong reasoning is described.
1	One of the following two is described.	One of the following two is described.	One of the following two is described.
	1. The circuit in (1) applies to the connection of the battery in (A).	1. The light bulb in the circuit in (1) is brighter than the light bulb in the circuit with one battery.	If you increase the number of batteries in series to two and turn on the light bulb, the current (voltage, power, and amount of electricity) flowing through the bulb will increase compared to when it is connected to a single battery.
	2. The circuit in (2) applies to the connection of the battery in (B).	2. The light bulb in the circuit in (2) is the same brightness as the light bulb in the circuit with one battery.	Even if you increase the number of batteries to two by connecting them in parallel and turn on the light bulb, the current (voltage, power, and amount of electricity) flowing through them does not change compared to when they are connected to a single battery.
2	Both of the above contents are described.	Both of the above contents are described.	Both of the above contents are described.

APPENDIX C

Table C. The rubric for the argument evaluation task

Case	Judgment	Reason
X	Incorrect	There are some points to the following effect that refer to insufficient (missing) reasoning. e.g. It is not written why we can say so from the results. There is no reason. The cause is not stated. It uses facts as reasons. Only facts are stated. The reason is a phenomenon. Does not explain the principle (law). Does not describe the nature (function) of a magnet. No theory.
Y	Incorrect	There are some points to the following effect that refer to insufficient (missing) claim. e.g. No answer to the question; does not say what poles A and B are, respectively. No conclusion. Only the result is stated. Only reasons are given.
Z	Incorrect	There are some points to the following effect that refer to insufficient (missing) evidence. e.g. The statement is not based on experimental results. No explanation of the phenomenon. Lack of facts.

https://www.ejmste.com