TÜRK FEN EĞİTİMİ DERGİSİ
Yıl 13, Sayı 4, Aralık 2016

Journal of
TURKISH SCIENCE EDUCATION
Volume 13, Issue 4, December 2016

http://www.tused.org

Examination of Science Learning Equity by Argumentation Instruction between Students Having Different Socio-Economic Status and Attending Different Achievement Level Schools

Ömer ACAR[1]

[1] Assis. Prof. Dr., Kocaeli University, Faculty of Education, Kocaeli-TURKEY

Received: 01.04.2016 **Revised:** 17.08.2016 **Accepted:** 06.09.2016

The original language of article is English (v.13, n.4, December 2016, pp.262-280, doi: 10.12973/tused.10184a)

ABSTRACT

We investigated if argumentation instruction provides equal science learning opportunities to students who have different socio-economic status (SES) and attend different achievement level schools. We selected a disadvantaged school and an advantaged school for this aim. 46 low-SES 8th graders in the disadvantaged school formed the experimental group. 35 low-SES 8th graders in the disadvantaged school formed the control group and 29 high-SES 8th graders in the advantaged school formed the comparison group. While experimental group received argumentation instruction on science topics during one semester, control and comparison groups did not. We compared group performances on conceptual knowledge, utility value of science, beliefs on theory and data, and views on student-centered teaching. Results showed that students of experimental and comparison groups outperformed students of control group on all measures after instruction. In addition, no difference was found between experimental and comparison group after instruction except from beliefs on theory and data measure.

Keywords: Achievement; Argumentation; Equity; Socio-economic status; Middle school.

INTRODUCTION

The importance of student evidence-based reasoning and reasoning between different alternative theories/explanations have been emphasized in science education and cognitive psychology research (Driver, Newton, & Osborne, 2000; Kuhn, Schauble, & Garcia-Mila, 1992; Zeidler, 1997). Studies mostly show that students have difficulty in evidence-based reasoning and reasoning between alternatives (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kelly, Druker, & Chen, 1998; Kuhn et al. 1992; Watson, Swain, & McRobbie, 2004; Zeidler, 1997). From this perspective, providing contexts to students where they can construct evidence-based arguments and reason between alternatives is recommended (Fleming, 1986; Kuhn, 1993).

Researchers have focused on teaching students how to construct a qualified argument and make a good argumentation in science classes. Encouraging results were obtained regarding the improvement of quantity and quality of evidence and these results warrant students use in their arguments (McNeill, Lizotte, & Krajcik, 2006; Sandoval & Milwood,

Corresponding author e-mail: acarok@gmail.com

© ISSN:1304-6020

2005). Moreover, it is demonstrated that students' arguments for other alternatives can be enhanced in science classrooms where students are provided with opportunities for counter-argument and rebuttal construction (Acar, 2008; Osborne, Erduran, & Simon, 2004a). In addition, studies showed that important variables in science education such as conceptual knowledge (Gültepe & Kılıç, 2013; McNeill et al., 2006; Zohar & Nemet, 2002), nature of science understanding (McDonald, 2010; Walker & Zeidler, 2007), and scientific reasoning (Acar, 2015) may be enhanced by fostering argumentation in science classes.

From the perspective of raising scientifically literate citizens, the focus of any student-centered instruction is changing from examination of effectiveness to achieving equity among students (National Research Council [NRC], 1996; Organisation for Economic Co-operation and Development [OECD], 2013). Recent studies have also focused on comparison of performances of low achieving students (LAS) and high achieving students (HAS) during inquiry instruction (Akkus, Gunel, & Hand, 2007; Lewis & Lewis, 2008). Encouraging results were obtained with regard to closure of achievement gap by gender and race (Geier et al., 2008; Huppert, Lomask, & Lazarowitz, 2002; Johnson, 2009; Wilson, Taylor, Kowalski, & Carlson, 2010). However, there is paucity of study examined this issue in the context of argumentation instruction. More specifically, Zohar and Dori (2003) compared reasoning and Akkus et al. (2007) compared achievement of LAS and HAS during argumentation instruction. These researchers sought whether argumentation instruction provided equal learning opportunities for both groups. However large-scale assessments show that socio-economic status (SES) and school type are also important variables for explaining students' science achievement (Martin, Mullis, Foy, & Stanco, 2012; OECD, 2013). Thus, researchers need to consider these variables for the examination of achieving equity among students in detail. Moreover, Acar (2015) compared conceptual knowledge and scientific reasoning of students in a disadvantaged school who received argumentation instruction with students in a comparatively advantaged school who received traditional instruction. In addition, a study by Chen, Hand, and McDowell (2013) examined the relative conceptual knowledge performances of low-SES students who received argumentation instruction and low-SES students who received traditional instruction. However, Acar (2015) did not pay attention to SES of students and Chen et al. (2013) did not compare the experimental group with high-SES students in an advantaged school. Furthermore, no research in argumentation literature have examined the equity among students for important variables in science education such as students' attitudes towards science, epistemological beliefs, and views on science teaching. This study aims to close these gaps in the literature. Following research questions were sought in this study:

Research Question 1: Is there any conceptual knowledge difference between control and experimental group low-SES students in disadvantaged school and comparison group high SES students in advantaged school after instruction?

Research Question 2: Is there any utility value of science and beliefs on theory and data difference between control and experimental group low-SES students in disadvantaged school, and comparison group high SES students in advantaged school after instruction?

Research Question 3: Is there any difference of views on student-centered teaching between control and experimental group low-SES students in disadvantaged school, and comparison group high SES students in advantaged school after instruction?

Argument and Argumentation

Argument and argumentation refer to different constructs in the literature. That is to say, argument is a product of one's attempt to support a claim about an issue (Kuhn, 1993; Kuhn & Udell, 2003). Thus, there need not to be alternative explanations or theories when one is constructing an argument. On the other hand, argumentation is a reasoning process by which advantages and disadvantages of alternative explanations or theories are examined (Kuhn 1993).

Toulmin (1958) offered a framework which can be used to construct model and assess arguments in practical situations (Toulmin, Rieke, & Janik, 1984). According to Toulmin (1958), a simple layout of an argument consists of data, warrant, backing, and claim. Data are the observations or facts that can be used to support a claim. A warrant is a reasoning that serves as a connection between data and the claim. A backing is a basic assumption in a domain that serves as a justification for the warrant. Finally, a claim is a conclusion stating one's stance on an issue. In more advanced arguments, qualifiers and rebuttals can also be used (Toulmin 1958). A qualifier is a statement that specifies the conditions under which the claim is true and a rebuttal is a statement that indicates the circumstances under which the claim is wrong.

Toulmin's argumentation pattern (TAP) was used both as an assessment technique for student arguments and as an instructional tool to teach evidence-based reasoning in science classrooms in the literature. Although several limitations of TAP was found for tracking student written and oral arguments in science classes (see Erduran, 2007 and Sampson & Clark 2008 for review), it has been widely used in science education literature to assess student arguments (e.g., Bell & Linn 2000; Jimenez-Aleixandre et al. 2000; Kelly et al. 1998, Watson et al. 2004). On the other hand, studies showed that explicating the components of TAP to students help students improve their written and oral arguments (Osborne et al., 2004a; Sandoval & Millwood, 2005; Zohar & Nemet, 2002).

As for argumentation, studies in cognitive psychology and science education show that subjects who are dependent on their theoretical beliefs demonstrate reasoning flaws when they argue among different alternative theories (Kuhn, 1991; Kuhn et al., 1988; Kuhn et al., 1992; Zeidler, 1997). Mostly they have difficulty in differentiation between theory and evidence (Kuhn, 1993). However, subjects who can offer evidence that is not theory oriented are more able to coordinate their theories with evidence (Kuhn, 1993; Kuhn et al., 1992). Accordingly, these latter subjects are more competent in arguing between different alternatives (Kuhn, 1991; Kuhn et al., 1988; Kuhn et al., 1992). Studies in science education also show that students mostly rely on their beliefs when they argue among alternative theories (Sadler, Chambers, & Zeidler, 2004; Zeidler, Walker, Ackett, & Simmons, 2002). In addition, they solely rely on scientific authorities without scrutinizing the data in their arguments when they argue between alternatives (Kolsto, 2001). As a remedy to these problems, providing students contexts where they can argue among different alternatives is recommended (Acar, 2008; Kuhn, 1993; Osborne et al., 2004a).

Effect of Argumentation Instructions on Student Related Variables

Studies show that students use rare evidence and justifications to support their claims in inquiry learning environments (Jimenez-Aleixandre et al., 2000; Kelly et al., 1998; Watson et al., 2004). Furthermore studies reveal that students are reluctant to consider other alternatives about an issue (Sadler et al., 2004; Zeidler, 1997; Zeidler et al., 2002). This is problematic in terms of inquiry learning because students are expected to construct and test evidence-based arguments in these settings (NRC, 1996). From this point of view, teaching argument and argumentation are seen as a remedy to these problems.

Several argumentation studies focused on if this kind of instruction helps students develop their argument and argumentation quality. Results demonstrate that student evidence and justification use can be enhanced through teaching the components of an argument (McNeill et al., 2006; Zohar & Nemet, 2002). Besides, it is found that students' skills related to arguing for the other alternatives can be enhanced (Acar, 2008; Nussbaum & Sinatra, 2003; Osborne et al., 2004a). Moreover, studies focus on the effect of argumentation instruction on important variables in science education such as conceptual knowledge, epistemological beliefs, nature of science, and attitudes towards science. Results of these studies suggest that students' conceptual knowledge and nature of science understanding may be enhanced through argumentation instruction (Aydeniz, Pabuccu, Cetin, & Kaya, 2012; Günel, Memiş, & Büyükkasap, 2010; McDonald, 2010; Zohar & Nemet, 2002). On the other hand, equivocal results are reported for the effect of argumentation instruction on students' epistemological beliefs (Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013; Sandoval & Morrison, 2003) and attitudes towards science (Osborne et al., 2013).

Achieving Equity in Science Classrooms and Argumentation Instruction

Science achievement gap among students has been a concern among policy makers (Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı, 2010; NRC, 1996; OECD, 2013). This gap can be attributed to the differences in student-level factors such as students' resources at home, self-efficacy, attitudes towards science, gender and their parents' SES and education level (Acar, Türkmen, & Bilgin, 2015; Engin-Demir, 2009; Martin et al., 2012; Sun, Bradley, & Akers, 2012; Yetişir, 2014); school-level factors such as program type, school climate (Dinçer & Uysal, 2010; Engin-Demir, 2009); classroom-level factors such as type of instructional method, class average engagement, and collaboration among teachers (Atar, 2014; Aypay, Erdoğan, & Sözer, 2007; Ceylan & Akerson, 2014; Yetişir, 2014). However, this issue is problematic with the movements of *science for all* and raising scientifically literate students who would participate in complex problems in the future as citizens (NRC, 1996).

Although addressing all of these factors in an education system for approaching achievement equity among students require large-scale policy efforts, several initiatives can be undertaken by researchers to reduce the achievement gap. Among one of these initiatives, studies examined if reform-based inquiry instruction leads to reduction of achievement gap in science classes. It is suggested that inquiry teaching is effective in eliminating the achievement gap by race (Johnson, 2009; Wilson et al., 2010), and gender (Geier et al., 2008; Huppert et al., 2002). On the other hand, it is also reported that traditional instruction results in a detectable achievement gap (Wilson et al., 2010). Contrary to these encouraging results, Lewis and Lewis (2008) found that students' prior Scholastic Aptitude Test scores explained a significant portion of their achievement in inquiry instruction which suggests that preexisting achievement gaps among students still did not close after inquiry. In another study, Von Secker and Lissitz (1999) examined data from a national study for the effect of instructional practices on eliminating the achievement gaps among students. More specifically, authors compared science achievement of several LAS groups based on gender, race, and SES in reform-based instruction. Authors found that achievement gap between LAS and HAS reduced in learning environments that focused on laboratory inquiry but widened in environments where critical thinking was fostered.

In case of argumentation instruction, four studies found in the literature which examined conceptual knowledge, reasoning, and achievement gaps between student groups during argumentation instruction. To begin with, Zohar and Dori (2003) examined thinking scores of high and middle school LAS and HAS during argumentation instruction. Actually, this study was an overall report of the four studies conducted previously. Students were grouped under LAS and HAS based on their previous science academic achievement. Mostly students in this study were required to develop their arguments with using evidence and consider alternative perspectives about contemporary science and society problems like bioethical dilemmas and diminishing of ozone layer. Overall, results suggested that both LAS and HAS developed their thinking skills. However, no consistent result was found for the closure of thinking scores between these groups. In another study, Akkus et al. (2007) investigated achievement of middle and high school LAS and HAS during argumentation instruction. Students' laboratory work was organized in this study in a way that students wrote their ideas, observations, claims, and reflections during the investigations. On the other hand, teachers monitored student learning during these processes. Researchers grouped students under achievement levels based on their performance on a baseline science test. Results showed that argumentation instruction helped to reduce the achievement gap between LAS and HAS in this study. In another study, Acar (2015) investigated the relative scientific reasoning and conceptual knowledge scores of students who were taught by argumentation instruction in a disadvantaged school and students who were taught by traditional instruction in an advantaged school. Competing theories and predict-observe-explain (POE) teaching strategies were used to foster student argument and argumentation during the study. Neither conceptual knowledge nor scientific reasoning difference was found between students in disadvantaged school and students in advantaged school. Finally, Chen et al. (2013) compared achievement of low-SES students who received argumentation instruction with other low-SES students who received traditional instruction. Students in the experimental group in this study wrote letters containing evidence-based arguments about force and motion to the questions posed by their older peers. Results showed that low-SES students in the experimental group outperformed low-SES students in the control group on force and motion achievement test.

As can be seen, Zohar and Dori (2003) and Akkus et al. (2007) examined whether argumentation instruction provides equal learning opportunities for LAS and HAS. However, for the investigation of science learning equity on a large scale, students' SES and their school type should be taken into account. Although Acar (2015) examined if students in a disadvantaged school taught by argumentation can close the scientific reasoning and conceptual knowledge gap with students in an advantaged school, he did not pay attention to students' SES. Furthermore, no control group from a disadvantaged school was included in that study. On the other hand, although Chen (2013) considered students' SES, this study did not compare low-SES and high- SES students and did not take school type into account. Addressing limitations of these studies is important for getting a clear picture for science learning equity. Therefore we aimed to address these gaps in the present study.

METHODOLOGY

Sample and Research Context

This research took place in an industrial city in Turkey in the spring academic semester. Quasi-experimental research design was utilized. More specifically, we selected a disadvantaged and an advantaged school for examining if argumentation instruction taught to students in the disadvantaged school helps to close the learning gap between these students and the students in the advantaged school. We selected these two schools based on their previous science achievement on a state-wide exam called passing from primary to middle school education exam (i.e., *Temel Eğitimden Ortaöğretime Geçiş sınavı* (TEOG) in Turkish). Specifically, disadvantaged and advantaged schools' means in science on this exam, that is used to place students in high schools and administered one semester before the study, were 58.16 and 78 out of 100 respectively. Families of the students' in the former school were mostly immigrants and had low-SES. On the other hand, families in the latter school mostly had high-SES. We focused our attention to 8th grade classes because we had developed argumentation activities and conceptual knowledge test for 8th grades in our previous work (see Acar, 2015). We did validity checks for both argumentation activities and conceptual knowledge test, and this enabled us to spend our effort on other methodological issues. Then we selected two 8th grade science classes as experimental and two 8th grade science classes as control group from disadvantaged school. Science teachers in this school who participated in this study said that four classes have same science achievement levels. In addition, an 8th

grade science class from the advantaged school was selected to compare other groups' performances. Although there were a total of 109 8th grade students in the disadvantaged school, 46 students remained in the experimental group, 35 students remained in the control group from the disadvantaged school and 29 students remained in the comparison group from the advantaged school after a list-wise deletion of four of the dependent variables used in this study. As can be seen, we had 28 students missing from the study sample in the disadvantaged school. Since we administered our instruments in two different days after TEOG exam after which student attendance rates usually get lower especially in disadvantaged schools, this high rate of missing data may be tolerable. However, the problem of whether missing students have different characteristics, e.g., initial conceptual knowledge, than the students remained in the study for disadvantaged school is a considerable one. To examine this problem, we performed ANOVA on students' conceptual knowledge pretest scores in the disadvantaged school. Result demonstrated that missing students (n = 28; M = 6.44) did not have different conceptual knowledge pretest scores than their peers (n = 81; M = 6.48) who remained in the study sample ($F_{(1, 107)} = 0.06$; $p > .05$). Therefore we do not have evidence for the claim that study sample is different than the students who were excluded from the sample.

We tested the assumptions related to student families' SES for each group and science achievement of control and experimental group in the disadvantaged school. To test the assumption regarding the SES of student families, we administered a questionnaire to students in which we asked education level of students' fathers and mothers, and monthly income of their families. We performed analysis of variance (ANOVA) on this overall SES measure. Result revealed that there is a significant difference between groups ($F_{(2, 107)} = 158.26$; $p < .001$). Then, we performed post-hoc comparisons with *Sidak* adjustment to experiment-wise alpha. Result showed that control (M = 6.40) and experimental groups (M = 6.15) in the disadvantaged school did not differ ($p > .05$). However, comparison group in the advantaged school (M = 10.03) had higher SES than control and experimental groups in the disadvantaged school ($p < .001$ for each comparison). Since the experimental and the control groups were in disadvantaged school, we expected similar science academic achievement between these groups. To test this hypothesis, we performed ANOVA on students' previous semester, i.e., fall semester, science academic grades. Result showed that experimental (M = 57.43) and the control group (M = 58.19) did not differ on their previous semester science grades ($F_{(1, 79)} = 0.06$; $p > .05$).

Three science teachers participated in this study. Two of them were in disadvantaged school and the other was in advantaged school. Although it might have been good for an experimental design to select one science teacher who could teach control and experimental groups as well as comparison group, it would be very difficult to find a science teacher who teaches at an advantaged and also disadvantaged school because of teacher appointment system in Turkey. However we could have selected two teachers one who had taught both control and experimental groups in disadvantaged school and the other who had taught comparison group in the advantaged school. On the other hand, Osborne et al. (2004) precaution other researchers that experimental group teachers can transfer their argumentation pedagogy to control groups unintentionally which may be undesirable for experimental design. In fact, our results regarding similar previous semester science achievement of control and experimental groups may confirm that control and experimental group teachers have similar pedagogical content knowledge because these teachers taught respective science classes in the previous academic semester. Experimental group's teacher attended to a professional development course on argumentation pedagogy before the study took place. To put it more clearly, components of TAP and argumentation were presented in the first session of this course. Additionally, examples of sound and fallacious argument and argumentation were presented and discussed in this session. Then, strategies to foster student argument and argumentation were presented in the second session. Finally, strategies for how to form small groups and how to scaffold students' argument and argumentation were presented and discussed in the final session. Each session lasted approximately half day period. In addition to the professional development course, the author gave feedback to this teacher about how to scaffold student argumentation in argumentation lessons better when he visited the class regularly throughout the semester. Control and comparison groups' science teachers taught the same science topics without argumentation.

The author visited and observed each group's science classes regularly during the semester. The author took field notes about the pedagogies teachers implemented during these observations. More clearly, the author paid particular attention to whether teachers used any argumentation strategies and any student-centered teaching approaches during these classroom observations.

Instruction

Experimental group's teacher explained TAP, argumentation and their components by examples in the first two class sessions of the spring semester. Then, experimental group did 6 argumentation activities during the semester and spent 6 lessons during these activities. There were sound, matter states and heat, living things and energy relation, electricity in our life and natural processes units covered in the spring semester of 8th grades. We used concept cartoons (Naylor & Keogh, 2013), competing theories, POE, six hats thinking techniques (de Bono, 1985) to develop argumentation activities. More specifically, two snowmen were presented as supporting alternative explanations about which one will melt first under sun shine in the concept cartoon activity (Osborne, Erduran, & Simon, 2004b) which was used in matter states and heat unit. Students first discussed this controversy in small groups and then wrote arguments on their worksheets. Competing theories strategy was used to develop two activities about how sound travels in a medium and how seasons form which were used in sound and natural processes units respectively. That is to say, two hypothetical students were presented for each topic as supporting alternative explanations. Another hypothetical student was also presented who provided data about the controversy. More clearly, a hypothetical student was presented as supporting the view that sound travels through the space of atoms and molecules of the medium in the first activity. Another hypothetical student was presented as arguing that sound waves moves by their effect on the particles of the medium. A third hypothetical student was presented as providing several everyday observations about the discussion such as sound waves are reflected as they encounter another medium and sound travels faster in the iron than it travels in the air. In the second activity, one of the hypothetical students was presented as claiming that seasons form due to the change of earth's distance from the sun. The other one was presented as claiming that the slope of the earth's orbit causes seasons. Finally, third hypothetical student was presented as showing data related to the controversy such as earth's having an elliptical orbit when rotating around the sun and when it is winter in northern hemisphere, there is summer in southern hemisphere. Students first discussed the controversy in small groups and then constructed their arguments, counter-arguments, and rebuttals in their worksheets for these two activities.

Two of the activities about the factors affecting the transformation of electric energy to heat energy and factors affecting the strength of an electro-magnet were developed using POE strategy which were used in electricity in our life unit. More clearly, students were asked to state dependent, independent, and controlling variables and predict which variables affect the water temperature in a beaker shown in Figure 1 for the first activity. Similar to the first activity, students were again asked to state dependent, independent, and controlling variables and predict which variables affect attraction force of electromagnet on clothespins. Then each group did the experiments and recorded the result of the investigations in their worksheets.

Finally, students explained their observations in the light of their predictions in their worksheets. Finally, six hats thinking technique was utilized to develop an argumentation activity which was about advantages and disadvantages about using nuclear energy which was used in living things and energy relation unit. More clearly, a scenario that explained the history of nuclear energy usage and evidence that is both for and against the nuclear energy usage was presented to students. For example nuclear power plant accidents in Three Mile Island and Chernobyl were presented as evidence against the usage of nuclear energy. On the other hand, scientists' view that nuclear power plants are safer for causing global warming than other power plants that use coal and petroleum which emit greenhouse gases was presented as evidence for supporting nuclear energy usage. First, each small group discussed this controversy in their groups. Then each student stated the facts (white hat), their emotions (red hat), advantages (yellow hat), disadvantages (black hat), innovative ideas (green hat), and finally their evaluation (blue hat) about the issue in their worksheets. Experimental group teacher guided student discussion in small groups. A transcript of a small group discussion can be seen in Table 1. As can be seen from this transcript, experimental group teacher fostered student argumentation in small groups for both sides of the controversy. Furthermore, he gave each student a chance in the small group to express his/her idea about the issue.

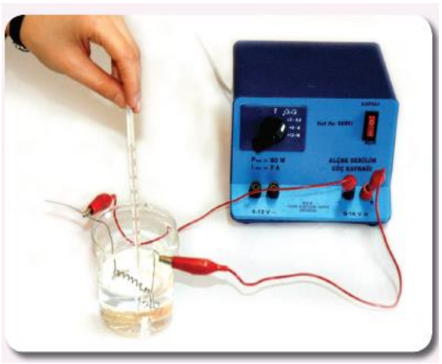


Figure 1. *Electric to Heat Energy Activity.*

Table 1. *Teacher-Student Interaction Transcript from an Argumentation Activity*

Teacher	Zeynep, what do you feel about (using) nuclear energy?
Zeynep	I think there are advantages and disadvantages (about using nuclear energy). I am really scared of its bad effects. If an accident happens, all of us can suffer.
Teacher	Ahmet (what do you feel)?
Ahmet	(a little pause) Nuclear energy is good for using (as an energy source) but it is bad for its negative effects (to environment).
Teacher	Ok, what about its advantages?
Sude	It is used for energy need.
Akın	It does not cause emission of greenhouse gases which contribute to global warming.
Teacher	Good, and what would you say about its disadvantages?
Zeynep	If an accident happens, people can suffer from it.
Ahmet	Furthermore, since its waste is dangerous for the environment, there is a problem with the waste disposal.
Teacher	All right, Akın do you have any innovative idea (about this issue)?
Akın	(hımmm) I think, (nuclear) power plants can be constructed far away from the cities.
Teacher	Sude (do you have any innovative idea)?
Sude	Its usage is good for our country if all the cautions can be undertaken (for an accident).
Teacher	Zeynep, can you give a summary of your stance about the issue?
Zeynep	I think nuclear energy is an important discovery because its positive effect on the economy. An accident can happen in a small probability if all the necessary cautions are undertaken. Therefore the probability of dangerous radiation coming from a nuclear power plant is as small as this probability.

Control and comparison groups' students were taught the same topics without argumentation. Male teacher in disadvantaged school taught science to control group mostly by lecturing. He did not even create a class environment for discussion during these occasions. In addition, he did not create a student-centered lesson that students can participate. Furthermore, he did not use the science laboratory for his lessons. On the other hand, female teacher in the advantaged school taught science to the comparison group by frequently using a laptop and a projector connected to it. Besides, she used smart board in each science lesson. Furthermore, she used science laboratory almost every week during the study. Thus, we cannot conclude that the latter group received traditional instruction because several elements of student-centered instruction were in place. Rather we can conclude that they did not receive argumentation instruction because these students neither argued between different alternatives nor constructed evidence-based arguments during the study.

Instruments

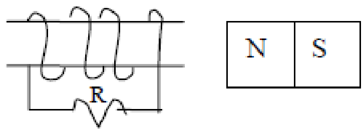
Conceptual Knowledge Test: This test was developed to measure 8^{th} graders' conceptual understanding related to sound, heat and temperature, states of matter and heat, electricity in our life, and natural processes. There were 17 multiple-choice items in the test. Content validity of the test for 8^{th} graders was established by Acar (2015). A student response was coded as 1 if he/she answered an item correct otherwise it was coded as 0. Cronbach's alpha estimate of the internal consistency was computed as .72 (n = 110) for the posttest administration of this test. Two example items of this test can be seen in Figure 2.

Matter	SpecificHeat (J/g °C)
Water	4,18
Lead	0,13
Iron	0,46
Olive oil	1,96

Same quantity of heat is given to the matters that have same temperature and mass initially and have specific heats indicated at the table. Which matter would have the highest temperature after the experiment is over?

a. Water b. Iron

c. Lead d. Olive oil



Which of the following situations cannot create electric current in the circuit shown on left?

a. Removing away the magnet from the circuit.
b. Bringing circuit closer to the magnet.
c. Moving the magnet and circuit in the same direction with the same velocity.
d. Rotating the magnet around N-S poles constantly.

Figure 2. *Two Example Items From Conceptual Knowledge Test.*

Utility Value of Science: Four Likert-type items were selected from PISA 2006 student questionnaire (OECD, 2006). These items were presented to a science education faculty to establish content validity. After examination of the questions, he stated that items were about a person's perception of science practical value in his/her daily life. Therefore, we named this questionnaire as utility value of science. Turkish translation of the questionnaire was done by the author and an English Language expert from Teaching English as a Second Language

Department edited any vague statement. All the items were in positive direction so if a student selected strongly agree, it was coded as 4 and if he selected strongly disagree, it was coded as 1. We computed Cronbach alpha as .55 (n = 110). Then we examined each items' contribution to the overall alpha. We recognized that an item was not contributing to the scale. After deletion of this item, we computed the alpha as .77 (n = 110). Accordingly, we performed statistical analyses with the remaining 3 items. Statements of these three items can be seen in Table 2.

Table 2. *Utility Value of Science Scale Items*

Advances in broad science and technology usually improve people's living conditions.
Broad science is important for helping us to understand the natural world.
I find that broad science helps me to understand the things around me.

Beliefs on Theory and Data: Leach, Millar, Ryder, & Sere (2000) developed this questionnaire to assess student epistemological beliefs about theory and data. There were 7 pairs of written statements which were about opposing philosophical stances on theory and data. Turkish translation of this questionnaire was done by the author and the English Language expert edited this translation. We coded each pair for representing contemporary view of theory and data or not. Then same categorization was done by a science education faculty. Both coding were consistent. For example "Scientists' theoretical assumptions influence their interpretation of data" was coded as representing contemporary view because it is related to theory-laden nature of science. On the other hand, "Scientists interpret data without being influenced by their theoretical assumptions" was coded as not representing contemporary view. Students were asked which statement they agree with for each pair of statements. If a student agreed with the contemporary view, his/her response was coded as 2 otherwise it was coded as 1. Then we examined internal consistency and found Cronbach's alpha as .52. Then we examined each item for their contribution to the scale. After this examination, we recognized that 3 items were not contributing to the scale. After deleting these items, Cronbach' alpha increased to .64 (n = 110). Thus, we used the remaining 4 items for statistical analyses. These statement pairs can be seen in Table 3.

Table 3. *Statements of Four Item Pairs in the Beliefs on Theory and Data Questionnaire*

The design of an experiment is dependent on theory about the thing that is being investigated.	An experiment is designed to see what happens, and does not depend on theory about the thing that is being investigated
Scientists interpret data without being influenced by their theoretical assumptions.	Scientists' theoretical assumptions influence their interpretation of data.
Scientists' ideas and theories influence their planning of data collection in experiments.	Scientists' put their ideas and theories to one side when they are planning data collection in experiments.
Scientists plan their data analysis based on the ideas and theories that they had when designing the experiment.	Scientists plan their data analysis without reference to the theories that they may have had when designing the experiment.

Views on Student-Centered Teaching: Four Likert-type items were selected from PISA 2006 student questionnaire (OECD, 2006) for examination of student views on student-centered teaching in their science classes. A science education faculty was also asked to examine items for content validity. He stated that items were assessing students' perception of science teaching in their science classes. We translated the items into Turkish and the English Language expert edited any vague statement. Meaning of each item was in positive direction.

Therefore we coded 4 for a response of in all lessons and 1 for a response of never or hardly ever. Internal consistency of Cronbach's alpha yielded to a score of .58 (n = 110). Each item in this questionnaire was contributing to the scale. Items in this scale can be seen in Table 4.

Table 4. *Items in the Views on Student-Centered Teaching Scale*

Students are given opportunities to explain their ideas.
Students spend time in the laboratory doing practical experiments.
Students are required to design how a school science question could be investigated in the laboratory.
The students are asked to apply a school science concept to everyday problems.

Data Analyses

For the first research question, we performed analysis of covariance (ANCOVA) on post conceptual knowledge scores of experimental and control groups. We checked normality, linearity of covariate over dependent variable and homogeneity of regression slopes assumptions for ANCOVA. Shapiro Wilk test showed post conceptual knowledge scores were normally distributed in experimental and control groups (W = .95, $p > .05$; W = .94, $p > .05$ respectively). Then we computed Pearson product-moment correlations between covariate and post conceptual knowledge scores. Result showed that there was a significant relationship between two variables (r = .24, $p < .05$). From this result, we concluded that linearity assumption was met. Finally we computed the F statistic for examining the homogeneity of regression slopes assumption. Result showed that this assumption was also met ($F_{(1, 77)} = 3.29$; $p > .05$).

Then, we performed ANOVA on post conceptual knowledge scores to examine any differences among the experimental, control, and comparison groups. We checked normality and homogeneity of variances assumptions. As reported previously, post conceptual knowledge scores were normally distributed in the former two groups. Result of the Shapiro Wilk test also showed this was the case for the comparison group (W = .94, $p > .05$). We examined Levene's statistic for homogeneity of variances assumption. Result showed that groups' variances for this variable were not same ($F_{(2, 107)} = 8.32$, $p < .001$). However, Lindman (1974) stated that F statistic is robust against the violation of homogeneity of variances assumption.

For the second and third research questions, we examined normality assumption for the remaining dependent variables. Results of the Shapiro Wilk tests showed that utility value of science scores were not normally distributed over the control, experimental, and comparison groups (W = .90, $p < .01$; W = .87, $p < .001$; W = .80, $p < .001$, respectively). Similar results were obtained for beliefs on theory and data scores (W = .78, $p < .001$; W = .85, $p < .001$; W = .81, $p < .001$, respectively). Finally, Shapiro-Wilk test showed while views on student centered teaching scores were normality distributed in comparison group (W = .95, $p > .05$), they were not normally distributed in control and experimental groups (W = .89, $p < .01$; W = .92, $p < .01$, respectively). Therefore, we used non-parametric tests to compare control, experimental, and comparison groups for these three dependent variables.

FINDINGS

Conceptual Knowledge Difference between Groups after Instruction

Means and standard deviations of pre and post conceptual knowledge scores of the three groups can be seen in Table 5. We performed ANCOVA for comparing post conceptual knowledge scores of low-SES students in the control group and low-SES students in the experimental group. Students' initial conceptual knowledge scores were the covariate in this analysis. Result of the ANCOVA showed that experimental group outperformed the control group controlling over initial conceptual knowledge scores ($F_{(1, 78)} = 35.21$; p < .001).

Table 5. *Descriptive Statistics of Groups' Conceptual Knowledge before and after Instruction*

		Pre Conceptual Knowledge		Post Conceptual Knowledge	
	n	M	SD	M	SD
Low-SES Students in Experimental Group	46	6.73	2.23	11.93	1.87
Low-SES Students in Control Group	35	6.17	1.20	8.40	3.24
High-SES Students in Comparison Group	29			12.31	3.01

Then we performed ANOVA to compare posttest conceptual knowledge scores of experimental, control, and comparison groups. The result showed that groups differed on this measure ($F_{(2,\ 107)} = 22.71$; $p < .001$). Post-hoc comparisons with Sidak adjustment showed that conceptual knowledge scores of low-SES students in the experimental group and high-SES students in the comparison group were similar ($p > .05$). On the other hand, high-SES students in the comparison group scored higher than low-SES students in the control group ($p < .001$).

Utility Value of Science and Beliefs on Theory and Data Difference between Groups after Instruction

For comparing groups' rank on utility value of science scores, we performed Kruskal-Wallis test. Then we did Mann-Whitney U-test to compare pairs of groups. Result of the Kruskal-Wallis test showed that groups differed on this measure, H = 65.85 (2, N = 110), p < .001. Mann-Whitney U-test's result showed high-SES students in the comparison group (Median = 11) outperformed low-SES students in the control group (Median = 7, U = 36.50, p < .001). In addition, low-SES students in the experimental group (Median = 10) outperformed low-SES students in the control group (U = 42, p < .001). However, no significant difference was observed between low-SES students in the experimental group and high-SES students in the comparison group (U = 508, p > .05).

Same analyses were performed for beliefs on theory and data scores. Kruskal-Wallis test result demonstrated groups' ranks were significantly different from each other H = 36.66 (2, N = 110), p < .001. Results related to Mann-Whitney U-tests showed low-SES students in the experimental group (Median = 6) scored higher than low-SES students in the control group (Median = 5, U = 471, p < .01), high-SES students in the comparison group (Median = 7) scored higher than low-SES students in the control group (U = 78, p < .001), and low-SES students in the experimental group (U = 359.5, p < .01).

Views on Student-Centered Teaching Difference between Groups after Instruction

Similarly we performed Kruskal-Wallis and then Mann-Whitney U-tests for views on student-centered teaching scores. Result of the Kruskal-Wallis test showed groups differed on this measure H = 18.57 (2, N = 110), p < .001. Mann-Whitney U-tests' results showed low-SES students in the experimental group (Median = 10) scored higher than low-SES students in the control group (Median = 8, U = 516, p < .01) but not high-SES students in the comparison group (Median = 12, U = 505.50, p > .05). Furthermore high-SES students in the comparison group scored higher than low-SES students in the control group (U = 191.50, p < .001).

To gain more insight on these differences, we performed separate Mann-Whitney U-tests on views on student-centered teaching items to compare low-SES students' views in the experimental group with that of low-SES students in the control and high-SES students in the comparison groups. Results showed low-SES students in the experimental group (Median = 4) scored higher than low-SES students in the control group on item 1 (Median = 2, U = 370.50, $p < .001$) and item 4 ($Median_{exp}$ = 3, $Median_{cont}$ = 2, U = 536.00, $p < .01$). More clearly, items 1 and 4 were about if students find chances to explain their ideas and if they find chances to apply school science to everyday problems in science classes respectively. However, no difference between these groups was found for item 2 ($Median_{exp}$ = 1.5, $Median_{cont}$ = 1, U = 728.50, $p > .05$) and 3 ($Median_{exp}$ = 1, $Median_{cont}$ = 1, U = 789.50, $p > .05$). Specifically, items 2 and 3 were about if students do practical experiments and if they investigate school science questions in their science labs respectively. Furthermore, results showed low-SES students in the experimental group (Median = 4) scored higher than high-SES students in the comparison group on item 1 (Median = 3, U = 497.00, $p < .05$). However, high-SES students in the comparison group outperformed low-SES students in the experimental group on item 2 ($Median_{exp}$ = 1.5, $Median_{comp}$ = 2, U = 412.50, $p < .01$) and 3 ($Median_{exp}$ = 1, $Median_{comp}$ = 2, U = 386.50, $p < .01$). On the other hand, no difference between these groups was observed for item 4 ($Median_{exp}$ = 3, $Median_{comp}$ = 3, U = 634.00, $p > .05$).

DISCUSSION

Our main aim in this study was to examine if argumentation instruction helps to achieve equity in science learning for different SES students attending low and high performing schools. For this aim, we formed experimental group consisting of low-SES students in a disadvantaged school who were taught science by argumentation, control group consisting of low-SES students in the same school and comparison group with high-SES students in an advantaged school who were not taught science by argumentation. Then we compared these groups' conceptual knowledge, utility value of science, views on theory and data, and views on student-centered teaching after instruction.

Our results suggest that low-SES students in the experimental group outperformed low-SES students in the control group on all of the measures. Furthermore, we did not find any difference between low-SES students in the experimental group and high-SES students in the comparison group on conceptual knowledge, utility value of science, and views on student-centered teaching total scores. Moreover, high-SES students in the comparison group outperformed low-SES students in the control group on all measures. These results are encouraging for researchers and educators who are concerned with equity issue among different student groups. More specifically, these results should be interpreted within the context of the gaps between different achievement schools because it is documented that students in high performing schools have more positive attitudes towards science and they understand science topics well conceptually than their peers in low performing schools (Aypay et al., 2007; Ceylan & Akerson, 2014). What our results suggest is then argumentation instruction may help in reducing the gaps between students who have low-SES and attend low achieving schools with their peers who have high-SES and attend high achieving schools. However, it should be kept in mind that method of instruction is only a factor, which we addressed, among others, i.e., student-level, classroom-level, and school-level factors, which contribute science achievement gap between student groups (Acar et al., 2015; Dinçer & Uysal, 2010; Martin et al., 2012). Nevertheless, the results are in alignment with previous research finding that instructional opportunities can compensate other factors that result in achievement gap (Von Secker, 2004; Von Secker & Lissitz, 1999). On the contrary, we found that traditional instruction resulted in science learning gap between low-SES students in low achieving school and high-SES students in high achieving school. Similar to these results, Acar (2015) found that the students who received argumentation instruction in a low achieving school closed conceptual knowledge and scientific reasoning gaps with their peers who received traditional instruction in a high achieving school. However, this study was limited in that it did not pay attention to students' SES. Furthermore, no control group consisting of low-SES students was included in that study. Chen et al. (2013) also found that low-SES students who received argumentation instruction had higher achievement scores than low-SES students who received traditional instruction. However, this study did not take school type into account and make a control group including high-SES students. Addressing these limitations is crucial for the examination of science learning equity because school type and students' SES are influential in explaining students' science achievement (Martin et al., 2012; OECD, 2013). We aimed to address these gaps in the literature. However, science learning gap cannot be limited to students who have different SES and attend schools which differ in science achievement. For instance, studies emphasized that other factors such as gender and race can cause science learning gap (Geier et al., 2008; Johnson, 2009; Wilson et al., 2010). Therefore, our results should be viewed as a part of a continuing endeavor to address science learning equity among different student groups.

Contrary to our expectation, high SES students in the comparison group outperformed low-SES students in the experimental group on beliefs on theory and data. An interpretation of this finding might be that development of low-SES students' beliefs on theory and data may require longer period of instruction. However, Osborne and his colleagues (2013) found no significant change of student epistemological beliefs after two year of argumentation instruction. Therefore, inspired by Sandoval and Morrison (2003), we suggest that nature of science issues should explicitly be taught to students in argumentation instruction to develop their epistemological beliefs.

When we analyzed students' views on student-centered teaching scores at item level, we found that low-SES students in the experimental group outperformed low-students in the control group on items 1 and 4 which were about if students find chances to explain their ideas and if they find chances to apply school science to everyday problems in science classes respectively. However, there was no difference between these groups on items 2 and 3 which were about science laboratory activities. Although students in the experimental group did two of the argumentation activities which were POE experiments, they did these activities in class but not in the lab. Therefore, this result is tolerable from this point of view. Low-SES students in the experimental group also outperformed high-SES students in the comparison group on item 1. This result also confirms that students in the experimental group had more opportunities than other groups for explaining their ideas in argumentation instruction. Although students in the experimental group had chances to apply school science to everyday problems such as how seasons form, and advantages and disadvantages of nuclear energy usage, there was not any difference between low-SES students in the experimental group and high-SES students in the comparison group on item 4. We interpret this result as a consequence of the science instruction that latter group received which had several elements of student-centered teaching. On the other hand, high-SES students in the comparison group outperformed low-SES students in the experimental group on items 2 and 3. This result is not surprising in that former group students frequently used their science lab during the instruction. However, latter group students had rare equipments and supplies in their science lab which was an obstacle for performing science classes in the lab. This result also raises an important problem for achieving equity among students. Achievement of equity would be possible if disadvantaged schools have equal science class and laboratory facilities and equipments with those in advantaged schools (Dincer & Uysal, 2010; Martin et al., 2012). One way to do this can be by clarifying class and laboratory standards nation-wide. Then each school needs can first be determined and then supplied by the use of these standards (Von Secker & Lissitz, 1999).

Limitations

There are several limitations in this study. First, our main aim in this study was to examine if argumentation instruction provides equal learning opportunities to students in a disadvantaged school with their peers in an advantaged school. However, our study may be limited in giving a full spectrum of examining of learning equity because we did not have an experimental group in the advantaged school. Thus, our results regarding learning equity should be interpreted within this limitation. Second, we used two science teachers in the disadvantaged school. We could have used a science teacher who teaches both control and experimental groups in the disadvantaged school to establish the experimental design in this study carefully. On the other hand, our finding regarding similar previous semester science achievement of control and experimental groups suggests that we do not have evidence that these two teachers were different in terms of their pedagogical content knowledge because same teachers had taught these groups in the previous semester before the study took place. Third, the period of professional development course given to the experimental group teacher may be viewed as relatively short according to longer periods of courses implemented in previous argumentation research (e.g., Erduran, Ardac, & Yakmaci-Guzel, 2006; Tümay and Köseoğlu, 2011; Simon, Erduran, & Osborne, 2006). However no time was devoted for training this teacher constructing materials and lesson plans for argumentation activities in the course because these activities were constructed by the author of this study. Therefore, this considerably shorter professional development course time should be evaluated within this context. In fact, there are pioneering studies in argumentation literature with having similar or shorter periods of teacher training and also have observed the effect of argumentation instruction (e.g., Akkus et al., 2007; Zohar & Nemet, 2002). Finally, we do not know each group's initial scores for utility value of science, beliefs on theory and data, and views on student-centered teaching. Therefore, we cannot claim any change for these variables during instruction. Future work is necessary which would assess each group's initial scores on these variables for examining the change from the beginning to the end of the instruction.

Acknowledgements

This research was supported by Kocaeli University Scientific Research Projects Department under research project number # HIDP2014/09.

REFERENCES

Acar, O. (2008). *Argumentation skills and conceptual knowledge of undergraduate students in a physics by inquiry class*. Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.

Acar, Ö (2015). Examination of science learning equity through argumentation and traditional instruction noting differences in socio-economic status. *Science Education International, 26*(1), 24-41.

Acar, Ö., Türkmen, L., & Bilgin, A. (2015). Examination of Gender Differences on Cognitive and Motivational Factors that Influence 8th Graders' Science Achievement in Turkey. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(5), 1027-1040. doi: 10.12973/eurasia.2015.1372a

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education, 29*(14), 1745-1765. doi: 10.1080/09500690601075629

Atar, H. Y. (2014). Multilevel effects of teacher characteristics on TIMSS 2011 science achievement. *Education and Science, 39*(172), 121-137.
Aydeniz, M., Pabuccu, A., Cetin, P. S., & Kaya, E. (2012). Argumentation and students' conceptual understanding of properties and behaviors of gases. *International Journal of Science and Mathematics Education, 10*(6), 1303-1324. doi: 10.1007/s10763-012-9336-1
Aypay, A., Erdoğan, M., & Sözer, M. A. (2007). Variation among schools on classroom practices in science based on TIMSS-1999 in Turkey. *Journal of Research in Science Teaching, 44*(10), 1417-1435. doi: 10.1002/tea.20202
Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*(8), 797-817. doi: 10.1080/095006900412284
Ceylan, E., & Akerson, V. (2014). Comparing the low- and high-performing schools based on the TIMSS in the United States. *Education and Science, 39*(173), 299-309.
Chen, Y. C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students' conceptual understanding: Learning about Force and motion through writing to older peers. *Science Education, 97*, 745-771. doi: 10.1002/sce.21067
de Bono, E. (1985). *Six thinking hats*. Boston, MA: Little, Brown.
Dincer, M. A., & Uysal G. (2010). The determinants of student achievement in Turkey. *International Journal of Educational Development, 30*, 592-598. doi: 10.1016/j.ijedudev.2010.05.005
Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287-312. doi: 10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A
Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı. (2010). *Uluslararası öğrenci değerlendirme programı PISA: 2009 ulusal ön raporu* [International student assessment program PISA: 2009 national preliminary report]. Ankara, Turkey: Earged.
Engin-Demir, C. (2009). Factors influencing the academic achievement of the Turkish urban poor. *International Journal of Educational Development, 29*, 17-29. doi: 10.1016/j.ijedudev.2008.03.003
Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education*, (pp. 47-69). Dordrecht, Netherlands: Springer.
Erduran, S., Ardac, D., & Yakmaci-Guzel, B. (2006). Learning to teach argumentation: Case studies of pre-service secondery science teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 2(2), 1-14.
Fleming, R. (1986). Adolescent reasoning in socio-scientific issues, Part II: Nonsocial cognition. *Journal of Research in Science Teaching, 23*, 689-698. doi: 10.1002/tea.3660230804
Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching, 45*(8), 922-939. doi: 10.1002/tea.20248
Gültepe, N. T., & Kılıç, Z. (2013). Bilimsel tartışma ve lise öğrencilerinin çözünürlük dengesi ve asitler-bazlar konularındaki kavramsal anlamaları. *Türk Fen Eğitimi Dergisi, 10*(4), 5-21.
Günel, M., Memiş, E. K., & Büyükkasap, E. (2010). Effects of the science writing heuristic approach on primary school students' science achievement and attitude toward science course. *Education & Science, 35*(155), 49-62.

Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education, 24*(8), 803-821. doi: 10.1080/09500690110049150

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*, 757- 792. doi: 10.1002/1098-237x(200011)84:6<757::aid-sce5>3.0.co;2-f

Johnson, C. C. (2009). An examination of effective practice: Moving toward elimination of achievement gaps in science. *Journal of Science Teacher Education, 20,* 287-306. doi: 10.1007/s10972-009-9134-y

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education, 20*(7), 849-871. doi: 10.1080/0950069980200707

Kolsto, S. D. (2001). 'To trust or not to trust,...'-pupils ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education, 23*(9), 877-901. doi: 10.1080/09500690010016102

Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press.

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

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327. doi: 10.1207/s1532690xci0904_1

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*(5), 1245-1260. doi: 10.1111/1467-8624.00605

Leach, J., Millar, R., Ryder, J., & Sere, M. G. (2000). Epistemological understanding in science learning: The consistency of representations across contexts. *Learning and Instruction, 10*, 497-527. doi: 10.1016/s0959-4752(00)00013-x

Lewis, S. E., & Lewis, J. E. (2008). Seeking effectiveness and equity in a large college chemistry course: An HLM investigation of peer-led guided inquiry. *Journal of Research in Science Teaching, 45*(7), 794-811. doi: 10.1002/tea.20254

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco, CA: W. H Freeman & Co.

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

McDonald, C. V. (2010). The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *Journal of Research in Science Teaching, 47*(9), 1137-1164. doi: 10.1002/tea.20377

McNeill, K. L., Lizotte, D. J., & Krajcik, J. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*(2), 153-191. doi: 10.1207/s15327809jls1502_1

National Research Council. (1996). *The national science education standards*. Washington, DC: National Academy Press.

Naylor, S., & Keogh, B. (2013). Concept cartoons: What have we learnt?. *Journal of Turkish Science Education, 10*(1), 3-11.

Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28,* 384-395. doi: 10.1016/s0361-476x(02)00038-3

Organisation for Economic Co-operation and Development. (2006). *OECD programme for international student assessment 2006: Student questionnaire*. Retrieved October 17, 2013, from http://pisa2006.acer.edu.au/downloads.php.

Organisation for Economic Co-operation and Development. (2013). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed* (Volume II). Paris, France: Author.

Osborne, J., Erduran, S., & Simon, S. (2004a). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020. doi: 10.1002/tea.20035

Osborne, J., Erduran, S., & Simon, S. (2004b). *Ideas, Evidence and Argument in Science* [In-service Training Pack, Resource Pack and Video]. London, England: Nuffield Foundation.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching, 50*(3), 315-347. doi: 10.1002/tea.21073

Sadler, T. D., Chambers, F. W., & Zeidler, D. L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education, 26*, 387-409. doi: 10.1080/0950069032000119456

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447-472. doi: 10.1002/sce.20276

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55. doi: 10.1207/s1532690xci2301_2

Sandoval, W. A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biology inquiry unit. *Journal of Research in Science Teaching, 40*(4), 369-392. doi: 10.1002/tea.10081

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education, 28*(2-3), 235-260.

Sun, L., Bradley, K. D., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Education, 34*(14), 2107-2125. doi: 10.1080/09500693.2012.708063

Toulmin, S. (1958). *The uses of argument*. New York, NY: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning* (2nd edition). New York, NY: Macmillan.

Tümay, H., & Köseoğlu, F. (2011). Kimya öğretmen adaylarının argümantasyon odaklı öğretim konusunda anlayışlarının geliştirilmesi. Türk Fen Eğitimi Dergisi, 8(3), 105-119.

Von Secker, C. (2004). Science achievement in social contexts: Analysis from national assessment of educational progress. *The Journal of Educational Research, 98*(2), 67-78. doi: 10.3200/joer.98.2.67-78

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching, 36*(10), 1110-1126. doi: 10.1002/(sici)1098-2736(199912)36:10<1110::aid-tea4>3.0.co;2-t

Walker, K. A., & Zeidler, D. L. (2007). Promoting discourse about socioscientific issues through scaffolded inquiry. *International Journal of Science Education, 29*(11), 1387-1410. doi: 10.1080/09500690601068095

Watson, J. R., Swain, J. R. L., & McRobbie, C. (2004). Students' discussions in practical scientific inquires. *International Journal of Science Education, 26*(1), 25-45. doi: 0950069032000072764

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching, 47*(3), 276-301. doi: 10.1002/tea.20329

Yetişir, M. İ. (2014). The multilevel effects of student and classroom factors on the science achievement of eight graders in Turkey. *Education and Science, 39*(172), 108-120.

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. *Science Education, 81*, 483-496. doi: 10.1002/(sici)1098-237x(199707)81:4<483::aid-sce7>3.0.co;2-8

Zeidler, D., Walker, K. A., Ackett, W. A., & Simmons, M. L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education, 86*, 343-367. doi: 10.1002/sce.10025

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences, 12*(2), 145-181. doi: 10.1207/s15327809jls1202_1

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. doi: 10.1002/tea.10008