Australian Journal of Teacher Education

Volume 41 | Issue 2 Article 5

2016

# Examination of Learning Equity among Prospective Science Teachers Who are Concrete, Formal and Postformal Reasoners after an Argumentation-Based Inquiry Course

Ömer Acar
*Kocaeli University*

Bruce R. Patton
*Ohio State University*



## Recommended Citation

Acar, Ö., & Patton, B. R. (2016). Examination of Learning Equity among Prospective Science Teachers Who are Concrete, Formal and Postformal Reasoners after an Argumentation-Based Inquiry Course. *Australian Journal of Teacher Education, 41*(2).
http://dx.doi.org/10.14221/ajte.2016v41n2.5

# Examination of Learning Equity among Prospective Science Teachers Who are Concrete, Formal and Postformal Reasoners after an Argumentation-Based Inquiry Course

Ömer Acar
Kocaeli University, College Of Education, Umuttepe Campus, Izmit,Turkey
Bruce R. Patton
Physics Department, The Ohio State University, Columbus, OH, USA

*Abstract: This study had two research purposes. First, we examined the scientific reasoning gains of prospective science teachers who are concrete, formal, and postformal reasoners in an argumentation-based physics inquiry instruction. Second, we sought conceptual knowledge and achievement gaps between these student groups before and after the instruction. Results were reported for 114 prospective science teachers. Results showed that concrete reasoners' scientific reasoning gain was higher than those of formal and postformal reasoners. Moreover postformal reasoners outperformed formal and concrete reasoners on a situational conceptual knowledge subscale before and after instruction. In addition, postformal and formal reasoners scored higher than concrete reasoners both on an initial achievement and final achievement measures. However, in-depth analyses showed that final achievement differences between postformal and concrete, and formal and concrete reasoners were lower than their respective initial achievement differences. Implications for teacher education programs were discussed according to these findings.*

## Introduction

Achieving equity in terms of student learning incomes and outcomes has been stressed as an important aim for science education in national and international guidelines (National Research Council [NRC], 2012; NGSS Lead States, 2013; The Organisation for Economic Co-operation and Development, 2013). From this perspective, research has examined if constructivist approaches to education help to achieve equity regarding student learning outcomes in science classrooms. More specifically, studies have compared the learning outcomes of low achieving students (LAS) and high achieving students (HAS) in both inquiry-based and traditional learning environments. The results demonstrate that students who received inquiry instruction outperformed their peers who received traditional instruction over several learning outcomes (Akkus, Gunel, & Hand, 2007; Dogru-Atay & Tekkaya, 2008; Geier et al., 2008; Huppert, Lomask, & Lazarowitz, 2002; Lewis & Lewis, 2008; Liao & She, 2009). Furthermore, inquiry teaching was found to be beneficial for historically disadvantaged students (Akkus et al., 2007; Geier et al., 2008; Wilson, Taylor, Kowalski, & Carlson, 2010). However, it is essential to examine the learning outcomes of different student groups within a classroom setting to ensure that any reform-based instruction creates equal learning opportunities for these students, which is a research recommendation that is part of "science for all" (NRC, 2012).

In a review of argumentation literature, we found a limited number of studies that examined learning gains of LAS and HAS in argumentation-based inquiry instruction. A study by Zohar and Dori (2003) aimed to compare the reasoning skills of middle and high school LAS and HAS during argumentation-based inquiry and traditional expository instruction. The authors categorized the students under LAS and HAS based on their previous science academic achievement. Findings showed that students in argumentation-based inquiry instruction gained higher reasoning skills than the students in traditional instruction. Moreover, it was found that both LAS and HAS benefited from argumentation-based inquiry instruction regarding reasoning skills. However little is known about the relative performances of LAS and HAS in scientific reasoning, conceptual knowledge, and achievement in this study. In addition, as argumentation is evidence-based reasoning, any result regarding this issue would be more meaningful if the performances of students with different reasoning levels were compared. Since the students' scientific reasoning skills were found to significantly predict student science achievement and conceptual knowledge in science classes (Ates & Cataloglu, 2007; Coletta & Phillips, 2005; Johnson & Lawson, 1998; Lawson, Banks, & Logvin, 2007; She & Liao, 2010), we think that students can be grouped under this variable to better analyze performance of students with different levels of reasoning ability in argumentation-based inquiry instruction.

Another neglected issue in argumentation literature is related to teacher education programs. Although argumentation intervention is integrated into teacher education programs in several studies (Acar, 2008, 2014; Zembal-Saul, 2009; Zembal-Saul, Munford, Crawford, Friedrichsen, & Land, 2002), no specific attention was paid to examine relative performances of students with differing levels of scientific reasoning. This issue is particularly important for prospective science teacher education programs because these teacher candidates will use the reasoning and argumentation skills developed during their education in their future as professionals. More research is needed in this domain to pinpoint the ways to improve the performance of prospective science teachers who are concrete reasoners. Therefore following research questions were examined in the present study:

R.Q.1: Do prospective science teachers with a low level of scientific reasoning enhance their scientific reasoning more than prospective science teachers with high level of scientific reasoning in an argumentation-based inquiry course?

R.Q.2: Do conceptual knowledge and achievement gaps decrease between prospective science teachers with different scientific reasoning abilities after an argumentation-based inquiry course?

## Conceptual and Theoretical Framework

Philosophers of science have emphasized the importance of argumentation involved in weighing and comparing different alternative theories for the development of science (Giere, 1984; Kuhn, 1996; Root-Bernstein, 1989). Hence the development of hypothetico-deductive reasoning is essential for students so they can select theories among rival theories and thus engage in high-quality scientific argumentation (Lawson, 2005, 2010).

Findings of both cognitive psychology and science education showed that subjects who adhere to their theoretical beliefs demonstrate reasoning flaws when they argue between different alternative theories. Mostly they have difficulty in coordinating their beliefs with evidence (Klaczynski, 2000; Kuhn, 2010; Kuhn, Iordanou, Pease, & Wirkala, 2008). However, subjects who can offer evidence that is not belief-oriented are more able to coordinate their theories with evidence. Accordingly, these latter subjects are more competent in arguing between different alternatives (Klaczynski, 2000; Kuhn, 1991; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn & Dean, 2004; Kuhn, Schauble, & Garcia-Mila, 1992). Studies in science education, on the other hand, have shown that students generally tend to rely on their beliefs when they argue between alternative theories (Acar, Turkmen, & Roychoudhury, 2010; Sadler, Chambers, & Zeidler, 2004; Zeidler, Walker, Ackett, & Simmons, 2002). In addition, students use wrong inclusion and exclusion of evidence in their arguments if they adhere to these theoretical beliefs (Kuhn et al., 1992). As a remedy to these problems, providing students contexts where they can argue between different alternatives using multiple sources of evidence is recommended (Acar, 2008, 2010; Kuhn, 2010; Osborne, Erduran, & Simon, 2004).

Students are expected to have control over their knowledge construction in inquiry learning environments with methods used by scientists (Abd-El-Khalick et al., 2004). More specifically, students are expected to engage in identifying problems, generating research questions, designing and conducting investigations, and formulating, communicating, and defending hypotheses and explanations in these contexts (Abd-El-Khalick et al., 2004). Similarly, according to a recent initiative for constructing a framework for K-12 science education, students are expected to engage in practices such as asking questions (for science) and defining problems (for engineering), developing and using models, planning and carrying out investigations, analyzing and interpreting data, using mathematics and computational thinking, constructing explanations (for science) and designing solutions (for engineering), engaging in argument from evidence, and obtaining, evaluating, and communicating information (NRC, 2012).

In essence, argumentation and inquiry are complementary structures in students' knowledge construction. That is, a student first needs to plan and carry out investigations, and then analyze and interpret data for preliminary steps in this process. Then he/she needs to construct evidence-based explanations, and counter-argue and critique other possible explanations for the selection of a more plausible explanation that interprets data best (Lawson, 2003, 2010; NRC, 2012). However, research has shown that student evidence-based reasoning in inquiry-based learning environments is problematic. Mostly, the students have difficulty with linking evidence and warrants to their claims (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kelly, Druker, & Chen, 1998; Watson, Swain, & McRobbie, 2004). As a remedy to this problematic evidence-based reasoning, several studies have incorporated argumentation teaching techniques into inquiry classes (e.g., Acar, 2008; Osborne et al., 2004; Zohar & Nemet, 2002). Encouraging results were obtained with regard to student argumentation and conceptual knowledge.

## Literature Review

### Achievement Gap in Inquiry and Argumentation Instruction

Experimental studies have shown the predominance of inquiry and argumentation teaching approaches in student learning over commonplace teaching (e.g., Geier et al., 2008; Wilson et al., 2010). However, efforts should go beyond from showing effectiveness to achieving equity among students of different abilities in inquiry classes (Lewis & Lewis, 2008). From this perspective, studies which focused on argumentation and inquiry compared learning outcomes of students with different achievement levels.

In the majority of the previous research, the learning outcomes of LAS and HAS in inquiry instruction have been examined at the middle school level (Geier et al., 2008; Johnson, 2009; Wilson et al., 2010). Additionally, a study by Akkus et al. (2007) examined the performance of LAS and HAS at the high school level and a study by Jackson and Ash (2012) examined the performance of the same student populations at the primary school level. The findings of these studies pointed out that race (Jackson & Ash, 2012; Johnson, 2009; Wilson et al., 2010) and gender (Geier et al., 2008) gaps were eliminated after inquiry instruction. In addition, Akkus et al. (2007) found that the achievement gap between LAS and HAS lessened after inquiry instruction.

Only one study by Lewis and Lewis (2008) investigated both the effect of inquiry instruction by forming a control group and comparing the learning outcomes of LAS and HAS in this inquiry instruction at the college level. Undergraduate students enrolled in a chemistry course were taught through peer-led guided inquiry in the experimental group and through lecture in the control group. Students in the experimental group worked in small groups and did activities which were led by a peer who was selected based on a good academic chemistry background. The guided inquiry used in this study was mostly based on the learning cycle teaching method. Results demonstrated that students in inquiry outperformed control group students regarding course achievement, which was measured by midterms and a final. Contrary to the expectation of the authors, findings pointed out that pre-existing achievement gaps among students did not lessen after inquiry instruction.

On the other hand, two studies were found in the literature which examined the learning performance of LAS and HAS in argumentation-based inquiry environments. Zohar and Dori (2003) examined the argumentation skills of high school students in an experimental group which received argumentation instruction and a control group which received traditional instruction. In addition the authors compared the argumentation skills of LAS and HAS in the experimental group. Findings showed that the experimental group students outperformed the control group students on argumentation skills. In addition, both LAS and HAS in the experimental group developed their argumentation skills during argumentation instruction. In another study, Acar (2014) categorized prospective science teachers into two groups, i.e., whether or not they had a consistent misconception about balanced forces. Acar (2014) found that there were scientific reasoning, conceptual knowledge, and achievement differences between these two student groups at the beginning of the instruction. However, after receiving argumentation-based inquiry instruction, the conceptual knowledge and achievement gaps between the groups were either closed or reduced.

In order to categorize students as LAS or HAS, Zohar and Dori (2003) referred to the students' science achievement background and Acar (2014) referred to whether the students had a consistent misconception or not. However in a science instruction that focuses on the development of reasoning skills as in the case of argumentation instruction, the categorization of students based on their scientific reasoning skills would give more reliable results. In fact Lawson (2010) states that argumentation and scientific reasoning are connected and a study by Schen (2007) demonstrates this connection. From this vein, it can be expected that students would develop their scientific reasoning in an argumentation-based instruction. However, the reviewed literature does not have a direct response to this hypothesis. In addition, a comparison of students with different scientific reasoning abilities in an argumentation-based inquiry course would show if this kind of instruction provides equal learning opportunities for students with low and high scientific reasoning levels. This research focus becomes more important when applied in science teacher education programs because little is known about the relative performances of prospective science teachers with different scientific reasoning levels in this kind of instruction. Examination of this research focus would reveal if argumentation-based inquiry instruction helps prospective science teachers who have a low level of scientific reasoning develop their science performance. Achieving equity among prospective science teachers is essential to ensure their qualifications as future education professionals (Acar, 2014).

Our perspective on achievement gaps among different student groups is in alignment with Lewis and Lewis (2008) in that it is possible to expect progress among both LAS and HAS in inquiry learning environments. However since HAS start any instruction with a substantial conceptual knowledge and reasoning background, it is fair to expect higher gains among LAS in inquiry settings, thus approaching equity in science classrooms.

#### Scientific Reasoning and Conceptual Knowledge

Lawson (1978) developed a test that can be used in classroom settings to identify students' formal reasoning level. A classroom test of formal reasoning was needed in science education research because administering each Piagetian task in classrooms was not efficient (Lawson, 1978). In early usages, this test was called as ‘formal reasoning’ test. There were items about control of variables, proportional, probabilistic, correlational, and combinatorial reasoning in the original version of the test. Subsequently items about hypothetico-deductive reasoning have since been included (Lawson et al., 2000). Recently this test has been referred to as the Classroom Test of Scientific Reasoning. In several studies, subjects were classified under scientific reasoning groups according to the scores they obtained from this test (Ates & Cataloglu, 2007; Lawson et al., 2007; Liao & She, 2009). To identify different scientific reasoners on objective grounds, Lawson (2003) established a set of guidelines for categorization. According to these guidelines, concrete reasoners are subjects who can seriate and classify objects, events, and situations; formal reasoners are the ones who can test causal operations using hypothetico-predictive reasoning; finally, postformal reasoners can test causal operations with unobservable entities using hypothetico-predictive reasoning.

Several studies examined the relation between students’ scientific reasoning skills and their misconception level. For instance, Acar (2014) categorized students under having a consistent misconception and those having a scientific conception based on their arguments about balanced forces. Acar (2014) then investigated scientific reasoning of these two groups. Acar (2014) found that students who had a misconception had lower scientific reasoning scores than their peers who had a scientific conception. In a pioneering study in this domain, Lawson and Worsnop (1992) analyzed the relation of high school students’ scientific reasoning skills with their misconceptions and their declarative knowledge about evolution. A negative correlation was found between students’ scientific reasoning abilities and misconception level. Furthermore, according to the results, students’ scientific reasoning levels predicted their declarative knowledge gain.

The association of scientific reasoning skills with pre- and post-instructional conceptual knowledge has been investigated in several studies. For instance, a study by Coletta and Phillips (2005) examined the relation between undergraduate students’ scientific reasoning and their conceptual knowledge gain related to Newtonian concepts. The authors found a strong positive relation between students’ scientific reasoning skills and their conceptual knowledge gains. Liao and She (2009), and She and Liao (2010) also found that 8$^{th}$ grader high scientific reasoners’ conceptual knowledge gains were higher than other 8$^{th}$ graders after a web-based learning unit. Similarly, Ates and Cataloglu (2007) investigated the relation of students’ scientific reasoning with their conceptual knowledge and problem-solving skills in an introductory mechanics course. A significant problem-solving difference among students with different reasoning abilities was detected. More clearly, postformal reasoners and formal reasoners outperformed concrete reasoners on this measure. On the other hand, no significant difference among reasoning groups was observed in pre- and post-test conceptual knowledge scores.

**Scientific Reasoning and Achievement**

Examination of the relation between students' scientific reasoning and their science achievement has been a research agenda in several studies. In a study by Johnson and Lawson (1998), the authors sought the effects of several scientific reasoning skills and prior biological conceptual knowledge on students' performance and achievement in expository and inquiry college biology classes. The results indicated that reasoning ability but not prior knowledge accounted for a significant amount of the variance on the students' final examinations. In addition, reasoning ability explained more of the variance on students' final examinations in expository instruction compared to inquiry instruction. In another study, Lawson et al. (2007) sought the relation between self-efficacy, scientific reasoning, and achievement in an introductory college biology course. Researchers found a positive significant correlation between scientific reasoning and self-efficacy. More importantly, scientific reasoning explained more of the variance in student achievement scores than self-efficacy. Similarly, She and Liao (2010) examined the relation of $8^{th}$ graders' scientific reasoning and conceptual knowledge with their achievement on a unit about atoms. Authors found that most of the variance in students' achievement was explained by their scientific reasoning scores.

## Method

**Research Design & Context**

Since we expected that both inquiry and argumentation approaches would help prospective science teachers achieve equity, we did not form a control group which received only argumentation or inquiry instruction. In addition, since a few selected physics topics were covered in this inquiry course, it would have been troublesome to form a control group which received instruction on the same physics topics by lecturing during this extended time. Instead we administered our instruments to a group of students receiving the same argumentation-based inquiry instruction. Thus our research design is a single group pretest-posttest design.

114 prospective science teachers enrolled in a Physics by Inquiry (PbI) course at a mid-western US university constituted the sample of this study. Most of these prospective science teachers were taking this course to fulfill their science credit requirement for graduation. Since PbI was offered as an introductory physics course, these students were taking the course before they specialized in any physics content areas. Of the participants whose data were included in the study, 74 of them were female and 40 students were male.

Since this sample size was too big for handling inquiry instruction, students were distributed to morning, afternoon, and evening sections. 40 students attended in the morning, 38 students attended in the afternoon, and 36 students attended the evening section. A multivariate analysis of variance was performed to examine if there were any pre-instructional scientific reasoning and conceptual knowledge differences among students in different sections. Result showed that students in different sections did not differ on the set of dependent variables (Wilks' $\Lambda$ was utilized; $F(6, 218) = 0.55$; $p > .05$). Follow-up analyses of variance also confirmed this finding for scientific reasoning and two subscales of conceptual knowledge, i.e., declarative and situational conceptual knowledge ($F(2, 111) = 1.05$; $p > .05$; $F(2, 111) = 0.70$; $p > .05$; $F(2, 111) = 0.36$; $p > .05$ respectively).

**Instruction**

Instruction lasted for 10 weeks. During this period, students met twice a week for a total of 6 hours per week. They worked in small groups consisting of three to four members. Students did experiments and exercises related to concepts of mass, balancing, volume, density, buoyancy, heat, and temperature in the Physics by Inquiry textbook volume 1 (McDermott, 1996). The small groups' reasoning and understanding were checked by instructors regularly. Instructional activities done at each class session can be seen in Tab. 1. The instructors gathered to discuss the ways to better scaffold student conceptual understanding and reasoning at these checks every week during the instructional period.

| Individual work | Group work | Teacher scaffolds |
|---|---|---|
| Students began each class with responding a question that is about the activities students did in the previous class session. | Each small group did the experiments and exercises in their textbook. Then each small group discussed about responses to the questions in their textbook | Instructors checked each small group's reasoning and conceptual understanding several times during a class session. |

**Table 1: Instructional activities during each class session**

| | Instruments | Instructional Activities |
|---|---|---|
| 1. week | Scientific reasoning pretest<br>Conceptual knowledge pretest | Guided inquiry: Examination of the effect of mass on balancing with using a balance and square nuts.<br>Argumentation: First written argumentation task about balancing and buoyancy. |
| 2. week | | Guided inquiry: Examination of the effect of the distance from the fulcrum on balancing using a balance and square nuts.<br>Argumentation: First oral argumentation task about balancing. |
| 3. week | First midterm | Guided inquiry: Examination of the effect of mass and volume on buoyancy<br>Argumentation: Second written argumentation task about balancing and buoyancy. |
| 4.-6. week | | Guided inquiry: Examination of the effect of objects' density on buoyancy.<br>Argumentation: Second oral argumentation task about buoyancy. |
| 7. week | Second midterm | Guided inquiry: Examination of the effect of liquids' density on buoyancy.<br>Argumentation: Third written argumentation task about balancing and buoyancy. |
| 8.-9. week | | Guided inquiry: Examination of algebraic expressions, graphs, and the relation and differences between heat and temperature. |
| 10. week | Scientific reasoning posttest<br>Conceptual knowledge posttest<br>Third midterm | Argumentation: Fourth written argumentation task about balancing and buoyancy. |

**Table 2: Sequence of the administration of instruments and instructional activities over the course period**

Both guided inquiry and argumentation teaching methods were utilized in PbI instruction. Sequence of the instructional activities related to guided inquiry and argumentation, and the administration of the instruments over the course period can be seen in Tab. 2. The learning cycle teaching method was used for guided inquiry. This teaching method has three phases: exploration, concept introduction, and concept application (Karplus, 1977). For instance, students in our study first did experiments using square nuts and a balance in the exploration phase to explore the relative effects of both mass and distance on moment. Then students were introduced to the concept of moment in the concept introduction phase. Finally they were required to apply the moment concept to a new situation in which the fulcrum was not in the middle in the concept application phase. The competing theories strategy (Bell & Linn, 2000; Osborne et al., 2004) was employed to construct four written and two oral argumentation tasks. Two hypothetical students were presented as supporting alternative explanations about balancing and buoyancy in these tasks. Everyday application examples of these concepts were also presented to students. Students were then asked to construct their arguments, counter-arguments (i.e., counter-arguing for the other alternative), and rebuttals (i.e., rebutting the other alternative). Students first discussed the hypothetical students' controversy and then constructed their arguments, counter-arguments, and rebuttals in small groups in oral argumentation tasks. Students first read the controversy presented in a work sheet for written argumentation tasks. Then they answered individually structured questions presented in this work sheet which fostered their arguments, counter-arguments, and rebuttals. An example of a written argumentation task can be seen in Fig. 1. Student learning and reasoning were checked by instructors after students finished both guided inquiry and argumentation tasks. No instruction occurred beyond these check points in the course. Instructors did not provide a direct feedback at these checks but rather guided student learning and reasoning by prompting questions. An excerpt transcribed from a check point after an oral argumentation task can be seen in Tab. 3.

| | |
|---|---|
| Student 1 | Observations a and b (a: bowl shaped clay floats in water whereas ball shaped clay with the same amount sinks in water, b: ship made of iron floats in water whereas a block of iron sinks in water.) would support student 1 (hypothetical student provided in student work sheets) |
| Instructor | Okay, why is that? |
| Student 1 | Because he is talking about how the shape, like a ship and like a ball shaped clay, in the same amount of the other that is made of same, because it not shaped in the same way. |
| Instructor | Okay, and student 1 is saying basically (intends to clarify student reasoning)? |
| Student 1 | Yeah that the shape of the object affects whether (thinks), like if it is bowl shaped it will float and if it is not it will sink |
| Instructor | Okay, student 2 is saying what? |
| Student 2 | The material… |
| Instructor | What do you mean by material? |
| Student 2 | Like what it is made of will affect whether it sinks or floats. |

**Table 3: Excerpt from buoyancy check point**

**Part 2. Sinking and Floating**

Two students were discussing sinking and floating. Below are their explanations:

**Student A:** "The mass of the objects affect whether they sink or float. Thus, more massive objects will sink whereas less massive objects will float in a given liquid."

**Student B:** "Both the mass and the volume of the object affect sinking and floating in a given liquid. Thus, if the mass is bigger and the volume is smaller, it will probably sink whereas if the mass is smaller and the volume is larger it will probably float."

Another friend provided some observations for their discussion. Those observations were:

**1.** A dry sponge floats high on the water but a water-soaked sponge will float level with the water surface.

**2.** An experiment showed that with careful pouring equal masses of 3 different liquids can form layers. It is observed that vegetable oil forms the top layer, water is in the middle layer, and corn syrup is on the bottom.

**3.** Both a small fish, like a goldfish, and a large fish, like a white shark, can float in the water.

**4.** A ship has a maximum load capacity. If the ship's load exceeds that amount, then the ship is in danger of sinking.

**Figure 1: Example of a written argumentation task (Acar, 2008; p. 145)**

**Instruments**

***Scientific Reasoning Test***

The Classroom Test of Scientific Reasoning was administered as a pre and posttest (see Tab. 2). This test was originally developed by Lawson (1978) to assess student formal reasoning skills such as conservation of mass, control of variables, proportional reasoning, correlational reasoning, probabilistic reasoning, and combinatorial reasoning. Additionally, questions related to hypothetical reasoning were added to the original version of the test in a study by Lawson et al. (2000). This revised version was used in the present study. This test comprises 12 two-tier multiple choice questions. Specifically, the first tier question is about a scientific reasoning skill and the second tier is about a justification to the first tier in each question set. Students' answers were coded as 1 if both the reasoning and justification questions were answered correctly; otherwise they were coded as 0. Cronbach's alpha estimate of internal consistency of the test was computed as .69 for the pretest and as .67 for the posttest ($n$ = 114).

Students were grouped into concrete, formal, and postformal reasoners according to their scientific reasoning pretest scores. Other studies have used several versions of the test depending on the suitability of these versions to their research aim. As a consequence, the number of questions and student scientific reasoning categorization differed slightly in these studies. For example, Lawson et al. (2007) used a version of the test with 11 two-tier questions for a total of 22 questions. The authors grouped the students into concrete reasoners if they scored between 0 and 9, formal reasoners if they scored between 10 and 18, and postformal reasoners if they scored between 19 and 22. In another study by Ates and Cataloglu (2007), the authors used a version of the test with 13 two-tier questions and categorized students based on their correct responses to two-tier question set. That is to say, students were grouped into concrete, formal, and postformal reasoners if they scored between 0 and 4, 5 and 9, and 10 and 13 respectively. The version with 12 two-tier questions used in a study by Coletta and Phillips (2005) was administered in the present study. Based upon the cutoff points used by Lawson et al. (2007) and Ates and Cataloglu (2007) and the prospective science teachers' score distribution on the scientific reasoning pretest in this study, students who scored between 0-5 were categorized as concrete reasoners; those who scored between 6-8 were grouped as formal reasoners; and those who scored between 9-12 were grouped as postformal reasoners. As a consequence, there were 30 students categorized as concrete, 51 as formal, and 33 as postformal reasoners.

#### *Conceptual Knowledge Test*

A 16-item multiple choice conceptual knowledge test was developed to assess student learning regarding the concepts taught in the course, i.e., mass, volume, density, balancing, uncertainty, buoyancy, interpretation of algebraic expressions and graphs, heat, and temperature. This test was administered as pre and posttest (see Tab. 2). Cronbach's alpha was computed as .47 ($n$ = 125) for the pretest and .55 ($n$ = 116) for the posttest.

A Principal Component Analysis (PCA) was performed on the posttest scores to examine any subscales. Another PCA for the pretest data was not performed because it was thought that student conceptual knowledge might have been fragmented at the pretest due to their unfamiliarity with the concepts before the instruction. Both eigen values and the scree plot were analyzed for the identification of the number of factors to be subtracted. Examination of eigen values showed 6 factors which had eigen values greater than 1. On the other hand, a closer look at the scree plot showed a big jump between the second and the third factor. Therefore two factors were selected for varimax rotation. In addition, factor loadings were suppressed to .3. Four items that had a loading less than .3 were removed from the analysis. Then Cronbach's alpha was computed for two subscales. After the examination of the item-factor correlations, one item that did not contribute to overall internal consistency of the first subscale was removed. Eventually Cronbach's alpha was computed as .60 for the first subscale consisting of 4 items and .47 for the second subscale consisting of 7 items. These two subscales explained the 27.24% variance of posttest scores.

The first author of this paper examined the items in each subscale, searching for any similar pattern between items. As a result of this process, it was discovered that the items in the first subscale were very similar to the exercises or questions students did in class. Although the items in the second subscale were indeed related to the concepts covered in the course, solutions to these items required a cognitive process of application of learning to novel situations. To establish the construct validity, the second author of this paper, who was also the principal instructor of the course, was asked to classify the items into recall and transfer questions. His classification of the items into recall and transfer questions was consistent with the results of the PCA excluding one item which was about heat and temperature. This item was identified as transfer in the PCA and as recall by the instructor. The authors held a discussion about any possibility of this item's possession of any transfer feature. The second author of this paper admitted that this item has also transfer features. As a conclusion, this item was included in the subscale which comprised transfer questions.

A study by de Jong and Ferguson-Hessler (1996) identified conceptual knowledge types. According to the authors, "declarative knowledge" includes recalling facts or formulas and "situational knowledge" includes the application of knowledge to novel situations. From this perspective, the first subscale was identified as declarative knowledge and the second subscale as situational knowledge. The items, their loadings, and the cognitive processes required to solve the items can be seen in Tab. 4 and Tab. 5. Item factor loadings, which can be seen in Tab. 4 and Tab. 5, were used to compute each conceptual knowledge type. As a result, a student could have a maximum score of 2.57 in declarative knowledge and a maximum score of 3.31 in situational knowledge. We did not make an equivalent scale, i.e., same maximum scores, for both subscales because we did not compare scientific reasoners' declarative knowledge with their situational knowledge. On the other hand, we examined scientific reasoners' declarative knowledge and situational knowledge gaps separately before and after instruction.

| Item | Loading | Knowledge | Cognitive process |
|---|---|---|---|
| 3 | .70 | Balancing | Applying $m_1 \times d_1 = m_2 \times d_2$ equation |
| 4 | .68 | Uncertainty | Finding the range of uncertainty |
| 5 | .67 | Conservation of mass | Recalling that mass conserves and volume can change |
| 7 | .52 | Volume | Applying m/d = v |

**Table 4: Items that loaded on declarative knowledge (Acar, 2008; p. 62)**

| Item | Loading | Knowledge | Cognitive process |
|---|---|---|---|
| 12 | .65 | Mass vs. volume graph and density | Using m/v for a heterogeneous object and interpretation of mass vs. volume graph |
| 11 | .58 | Sinking & floating and density | Reasoning involves sinking and floating behavior of a heterogeneous object will depend on density of its component objects |
| 15 | .52 | Heat and temperature | Contrast of 1g vs. whole object's heat and temperature by applying heat and temperature knowledge |
| 2 | .42 | Conservation of mass | Application of conservation of mass knowledge to a place where gravity is different |
| 1 | .42 | Balancing | Application of moment knowledge to a seesaw where fulcrum is not in the middle |
| 13 | .40 | Volume, mass | Interpretation of volume vs. mass graph using mass and volume knowledge |
| 10 | .32 | Sinking & floating and density | Reasoning that sinking and floating behavior of two objects will depend on objects' and liquids' densities |

**Table 5: Items that loaded on situational knowledge (Acar, 2008; p. 63)**

***Achievement***

Students' first midterm and final grades were the initial and final achievement measures. The first midterm included conceptual questions regarding the concepts of mass, balancing, volume, and density. It was administered in the third week of the course (see Tab. 2). Students' final grade was a weighted average of the course's three midterm exams and student assignments. Student assignments included homework, journal entries and question of the day. For each of the 10 weeks of the instructional period, the students answered questions about the concepts they had learned in the previous week in the homework assignment. Students reflected in their journals four times during the course about their opinion of their learning. The question of the day assignment was administered for each class session and reviewed the concepts students learned in previous class sessions. Each midterm and student assignment was constructed by the second author. In addition these achievement measures were reviewed by other instructors of the course for content validity.

#### Statistical Analyses

Analyses, dependent, and independent variables related to each research question can be seen in Tab. 6. For the first research question, we first performed separate paired *t* tests for each scientific reasoning group to examine their scientific reasoning change from pre- to posttest. Second, we performed an analysis of variance (ANOVA) on their scientific reasoning gains. First we examined normality assumption for this analysis. Results of Shapiro-Wilk tests showed scientific reasoning gains were normally distributed over concrete, formal, and postformal reasoners ($W = .97$, $p > .05$; $W = .96$, $p > .05$; $W = .95$, $p > .05$ respectively). Second we examined if the data violates the homogeneity of variances assumption. The result of the Levene test showed the reasoning gain variances among reasoners were similar ($F$ (2, 111) = 2.97, $p > .05$). Finally we performed pair-wise comparisons. We adjusted the experiment-wise alpha level to .05 using the Bonferroni correction in these comparisons.

For the second research question, we first aimed to reveal any initial conceptual knowledge and achievement gap among the reasoners. Then we investigated if these gaps closed or diminished after instruction. For the first aim, we performed a multivariate analysis of variance (MANOVA), which takes into account the relation of dependent variables, on two pretest conceptual knowledge subscales. We examined the Box test for the equality of covariances assumption for MANOVA and found that the covariances are equal ($F = 1.49$; $p > .05$). Then to pinpoint any significance, we first performed follow-up ANOVAs and then pair-wise comparisons with the Bonferroni correction. After an examination of the reasoners' pretest conceptual knowledge measures, we ran an ANOVA on the students' first midterm grades. First we examined normality assumption for this analysis. Results of Shapiro-Wilk tests showed that normality assumption was met for concrete, formal, and postformal reasoners ($W = .94$, $p > .05$; $W = .95$, $p > .05$; $W = .95$, $p > .05$ respectively). Second we examined homogeneity of variances assumption. The result of the Levene test yielded a significant score which meant that variances among reasoners were not similar in first midterm grades ($F$ (2, 111) = 7.85, $p < .005$). Although the $F$ test is quite robust regarding violations of the homogeneity of variances assumption, the actual alpha level would have been inflated. However our results yielded significance values lower than .005 which we thought may address this problem. Then we performed pair-wise comparisons with the Bonferroni correction.

For the second aim in the second research question, we ran two separate ANOVAs, one for posttest situational conceptual knowledge and one for the students' final grades. First we examined normality assumption for these analyses. Results of Shapiro-Wilk tests showed that normality assumption was met for concrete, formal, and postformal reasoners' posttest situational conceptual knowledge ($W = .96$, $p > .05$; $W = .98$, $p > .05$; $W = .96$, $p > .05$ respectively). Similar results were found for concrete, formal, and postformal reasoners' final grades ($W = .94$, $p > .05$; $W = .95$, $p > .05$; $W = .96$, $p > .05$ respectively). Second we examined homogeneity of variances assumption. Levene's test results for posttest situational conceptual knowledge and final grades showed the variances among the reasoners were similar ($F$ (2, 111) = 0.54, $p > .05$; $F$ (2, 111) = 1.30, $p > .05$ respectively). Then we performed pair-wise comparisons with the Bonferroni correction for each ANOVA. Finally we performed a repeated measures MANOVA on both situational conceptual knowledge and achievement measures to examine if the group differences in the pretest were similar to or different than the group differences in the posttest. Testing time, i.e., pretest and posttest, was the within-subjects factor and reasoning level was the between-subjects factor in these analyses. We examined the Box test for equality of covariances assumption and found that the covariances are equal for situational conceptual knowledge ($F = 1.21$; $p > .05$) but not for achievement measures ($F = 4.24$; $p < .005$) in these analyses. Although violation of this assumption for achievement measures may have inflated the actual alpha level, our results regarding achievement measures yielded significance values below the .001 level which we thought may compensate this violation. Finally, we ran interaction contrasts between scientific reasoning groups.

| | Part | Analyses | Dependent variable | Independent variable |
|---|---|---|---|---|
| 1. Research question | 1 | Paired t tests | Scientific reasoning pretest and posttest scores | ---- |
| | 2 | 1. ANOVA<br>2. Pair-wise comparisons | Scientific reasoning gains | Scientific reasoning groups |
| 2. Research question | 1 | 1. MANOVA<br>2. Follow-up ANOVA<br>3. Pair-wise comparisons | Situational & declarative conceptual knowledge pretest scores | Scientific reasoning groups |
| | | 4. ANOVA<br>5. Pair-wise comparisons | First midterm | Scientific reasoning groups |
| | 2 | 1. ANOVA<br>2. Pair-wise comparisons | Posttest situational conceptual knowledge scores | Scientific reasoning groups |
| | | 3. ANOVA<br>4. Pair-wise comparisons | Final grades | Scientific reasoning groups |
| | 3 | 1. Repeated measures MANOVA<br>2. Interaction contrasts | Pretest-posttest situational conceptual knowledge scores | Within-subjects factor: Testing time<br>Between-subjects factor: Scientific reasoning groups |
| | 4 | 1. Repeated measures MANOVA<br>2. Interaction contrasts | First midterm-final grades | Within-subjects factor: Testing time<br>Between-subjects factor: Scientific reasoning groups |

**Table 6: Description of the analyses performed for each research question**

## Results

### Scientific Reasoning Change

Descriptive statistics were computed for concrete, formal, and postformal reasoners' pretest and posttest scientific reasoning scores (see Tab. 7). To examine the change from pretest to posttest, paired *t* tests were performed for each group of scientific reasoners. Results showed that both concrete and formal reasoners increased their scientific reasoning scores during the instruction ($t(29) = 6.01$; $p < .05$; $t(50) = 4.15$; $p < .05$, respectively). Furthermore, concrete reasoners' increase had a large effect (Cohen's $d = 1.01$) and formal reasoners' increase had a medium effect (Cohen's $d = 0.58$) according to Cohen's rule for effect sizes (1988). However, postformal scientific reasoners' score did not increase ($t(32) = 0.67$; $p > .05$).

| | | Pretest | | Posttest | |
|---|---|---|---|---|---|
| | *N* | *M* | *SD* | *M* | *SD* |
| Concrete | 30 | 3.80 | 1.13 | 6.07 | 2.42 |
| Formal | 51 | 7.06 | .73 | 7.96 | 1.56 |
| Postformal | 33 | 9.88 | .99 | 10.03 | 1.38 |

**Table 7: Scientific reasoners' pretest and posttest scientific reasoning statistics**

ANOVA was performed on the scientific reasoning gains data. In this analysis, the scientific reasoning level was the independent variable and the scientific reasoning gain was the dependent variable. Result showed that scientific reasoners differed significantly in their gains ($F$ (2, 111) = 13.41, $p < .001$). Moreover, this difference had a medium practical significance ($\eta^2 = .20$). Post-hoc comparisons with the Bonferroni correction of the experiment-wise alpha level to .05 showed concrete reasoners' scientific reasoning gains ($M_{gain} = 2.27$) were higher than that of formal ($M_{gain} = 0.90$; $p < .01$) and postformal reasoners ($M_{gain} = 0.15$; $p < .001$). However, the formal reasoners' gains were not higher than postformal reasoners' ($p > .05$). The result of the comparison between concrete and formal reasoners had a medium practical significance (Cohen's $d = 0.75$), and between concrete and postformal reasoners had a large practical significance (Cohen's $d = 1.23$).

#### Conceptual Knowledge and Achievement Gaps

##### *Gaps Before Instruction*

Concrete, formal, and postformal scientific reasoners' pretest and posttest mean and standard deviation scores of declarative and situational knowledge and achievement can be seen in Tab. 8. First, analyses were performed for pretest measures for the investigation of conceptual knowledge and achievement differences among reasoners before instruction. Since both declarative and situational knowledge are conceptual knowledge constructs, a MANOVA test, which takes into account the relation of dependent variables, was run on the pretest conceptual knowledge subscales. A significant effect of reasoning level was obtained on the set of dependent variables (*Wilks' Λ* was utilized; $F$ (4, 220) = 4.40; $p < .005$). An examination of effect size showed a small practical significance of this result ($\eta^2 = .07$). Follow-up ANOVA results showed a significant effect of reasoning level on situational knowledge ($F$ (2, 111) = 8.32; $p < .001$) but not on declarative knowledge ($F$ (2, 111) = 1.95; $p > .05$). Furthermore, the situational knowledge difference among reasoners had a medium practical significance ($\eta^2 = .13$). Pair-wise comparisons with the Bonferroni correction showed postformal reasoners' situational knowledge ($M = 1.23$) was higher than formal ($M = 0.78$, $p < .01$) and concrete reasoners ($M = 0.60$, $p < .001$). Examination of the effect sizes showed the difference between postformal and formal reasoners had a medium significance and the difference between postformal and concrete reasoners had a large practical significance (Cohen's $d = 0.65$; Cohen's $d = 1.00$ respectively). On the other hand, the other comparison result showed formal and concrete reasoners' situational knowledge scores were similar ($p > .05$).

| | Concrete reasoners pretest | | Concrete reasoners posttest | | Formal reasoners pretest | | Formal reasoners posttest | | Postformal reasoners pretest | | Postformal reasoners posttest | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Declarative knowledge | .57 | .54 | 2.11 | .68 | .69 | .52 | 2.37 | .45 | .84 | .56 | 2.46 | .36 |
| Situational knowledge | .60 | .49 | 1.27 | .72 | .78 | .63 | 1.66 | .81 | 1.23 | .74 | 2.20 | .68 |
| Achievement [a] | 79.03 | 12.55 | 89.46 | 4.47 | 88.35 | 7.79 | 92.96 | 3.36 | 92.27 | 5.40 | 94.05 | 3.37 |

[a] The first midterm and the final grades were pretest and posttest achievement measures respectively.

**Table 8: Scientific Reasoners' Pretest and Posttest Descriptive Statistics of Conceptual Knowledge and Achievement**

An ANOVA was performed to examine any initial achievement gap among reasoners. In this analysis, reasoning level was the independent variable and the first midterm grade was the dependent variable. There was a significant effect of reasoning level on students' first midterm grades ($F$ (2, 111) = 18.96; $p < .001$). In addition this effect had a large practical significance ($\eta^2 = .26$). Postformal ($M = 92.27$) and formal reasoners ($M = 88.35$) had higher midterm grades than concrete reasoners ($M = 79.03$) according to the results of post-hoc comparisons with the Bonferroni correction (for each comparison $p < .001$). Examination of the effect sizes revealed that achievement differences between postformal and concrete reasoners, and formal and concrete reasoners both had large practical significances (Cohen's $d = 1.37$; Cohen's $d = 0.89$, respectively). No significance was detected for the comparison of postformal and formal reasoners' first midterm grades ($p > .05$).

#### *Gaps After Instruction*

To examine if initial situational knowledge and achievement gaps close among concrete, formal, and postformal reasoners after instruction, analyses were performed on student situational knowledge posttest scores and final grades. First an ANOVA was performed on posttest situational knowledge scores. Result pointed out a significant effect of reasoning level ($F$ (2, 111) = 12.17; $p < .001$). Moreover, this result had a medium practical significance ($\eta^2 = .18$). Post-hoc comparisons with the Bonferroni correction pinpointed this significance. According to the results, postformal reasoners ($M = 2.20$) scored higher than formal ($M = 1.66$, $p < .01$) and concrete reasoners ($M = 1.27$, $p < .001$). The other comparison did not reveal any significance ($p > .05$). According to Cohen's rule (1988), the situational knowledge difference between postformal and concrete reasoners had a large practical significance (Cohen's $d = 1.33$) and the difference between postformal and formal reasoners had a medium practical significance (Cohen's $d = 0.72$).

To examine if the posttest and pretest situational knowledge gaps between groups are similar or different, a MANOVA with repeated measures was performed. Testing time, i.e., pretest and posttest, was the within-subjects factor and reasoning level was the between-subjects factor in this analysis. According to the result, the interaction effect between time and reasoning level was not significant ($F$ (2, 111) = 0.94; $p > .05$). Besides interaction contrasts, i.e., comparing the differences of groups at the pretest with that of at the posttest, between postformal and formal reasoners ($F$ (1, 111) = 0.25; $p > .05$), and postformal and concrete reasoners ($F$ (1, 111) = 1.81; $p > .05$) did not reveal any significance which means that the group differences on the pretest were similar to the group differences on the posttest.

A second ANOVA was performed on student final grades to examine if there was an achievement gap between groups at the end of the instruction. According to the result, reasoning level had a significant effect on final grades ($F$ (2, 111) = 13.40; $p < .005$). The effect size showed a medium practical significance of this result ($\eta^2 = .19$). To pinpoint this significance, post-hoc comparisons with the Bonferroni correction were performed. According to these analyses, postformal ($M$ = 94.05) and formal reasoners ($M$ = 92.96) outperformed concrete reasoners ($M$ = 89.46, $p < .001$ for each comparison) on final grades. The effect sizes showed both comparisons of postformal and concrete, and formal and concrete reasoners had large practical significance (Cohen's $d$ = 1.16; Cohen's $d$ = 0.89 respectively). The other comparison did not reveal any significance ($p > .05$).

The interaction effect between testing time and reasoning level was scrutinized. A MANOVA with repeated measures was run on achievement measures, i.e., the first midterm and final grades. The result showed a significant interaction effect ($F$ (2, 111) = 12.22; $p < .001$). Eta squared showed this result had a medium practical significance ($\eta^2 = .18$). For in-depth analysis, an interaction contrast between postformal and concrete reasoners was performed. This analysis revealed that the gap between these groups in the first midterm was not the same as the gap in the final grades ($F$ (1, 111) = 23.50; $p < .001$). Examination of the effect size showed this result had a medium effect ($\eta^2 = .18$). According to the descriptive statistics given in Tab. 8, this result means that the achievement gap between these groups in the final grade was statistically lower than the gap in the first midterm. A second interaction contrast between formal and concrete reasoners was scrutinized. This analysis also revealed a significant result ($F$ (1, 111) = 12.82; $p < .001$) meaning the achievement gap between formal and concrete reasoners in the final grade was statistically lower than the gap between these groups in the first midterm. This significance had a medium effect ($\eta^2 = .10$). On the other hand, the other interaction contrast between postformal and formal reasoners did not reveal a significance ($F$ (1, 111) = 3.19; $p > .05$).

## Discussion

This study had two research purposes. First we examined if scientific reasoning gain of prospective science teachers who are concrete reasoners was higher than that of prospective science teachers who are formal and postformal reasoners in an argumentation-based inquiry course. Second, we examined if conceptual knowledge and achievement differences between prospective science teachers who have different scientific reasoning levels decrease after an argumentation-based inquiry instruction.

Results regarding the first research question showed only concrete and formal reasoners enhanced their scientific reasoning during the instruction. Examination of the effect sizes showed that concrete reasoners' scientific reasoning development had large practical significance and formal reasoners' development had medium practical significance. In addition, concrete reasoners' scientific reasoning gains were higher than those of formal and postformal reasoners with medium and large effect sizes respectively. Although previous research has shown that it is possible to enhance student scientific reasoning (e.g., Gerber, Cavallo, & Marek, 2001; Johnson & Lawson, 1998; Lawson et al., 2007; Marušić & Sliško, 2012) and achieve equity among different scientific reasoners in inquiry classes (Jensen & Lawson, 2011), little was known about whether scientific reasoning gaps between prospective science teachers who are concrete, formal, and postformal reasoners can be lessened in inquiry classroom settings. More specifically, studies showed that students enhanced their scientific reasoning in learning environments in which they were fostered to construct evidence-based explanations (Lawson et al., 2007; Marušić & Sliško, 2012). Similarly, prospective science teachers' scientific reasoning gain in the present study was not surprising in that they were also fostered to construct evidence-based explanations in this argumentation-based inquiry course. In addition to this scientific reasoning gain, the results of the present study also show that scientific reasoning gaps between low and high scientific reasoning prospective science teachers can indeed be reduced in argumentation-based inquiry classroom environments. This result is encouraging in the context of teacher education programs because it demonstrates that it is possible to achieve scientific reasoning equity among prospective science teachers who will scaffold their students' reasoning in the future as professionals.

Results regarding the second research question show that situational knowledge and achievement gaps, which were in favor of high scientific reasoners, occurred among reasoners at the beginning of the instruction. More specifically, postformal scientific reasoners outperformed formal and concrete scientific reasoners on a situational knowledge subscale with medium and large effect sizes respectively. Moreover, postformal and formal scientific reasoners scored higher than concrete scientific reasoners on the first midterm with both comparisons having large effect sizes. These findings are not new to the literature in that previous research has also indicated that good scientific reasoners have high conceptual knowledge and achievement (Coletta & Phillips, 2005; Johnson & Lawson, 1998; Lawson & Weser, 1990; Liao & She, 2009). What is novel in this research is that the findings shed light on which conceptual knowledge type made a difference among students with different scientific reasoning levels. According to the results, there was not any gap among the groups regarding declarative knowledge, i.e., conceptual knowledge related to recalling facts or formulas. However, scientific reasoners differed in situational knowledge, which is the knowledge related to the application of learning to novel situations. From this result it can be implied that one's situational conceptual knowledge ecology is related to his/her scientific reasoning level.

Investigation of posttest measures indicates that situational knowledge and achievement gaps between groups before the instruction still existed after the instruction. Similar results were obtained by Johnson and Lawson (1998), and Liao and She (2009) since these studies also showed that scientific reasoning level still explained student achievement after an inquiry instruction. On the other hand, the results of the interaction effect between testing time and reasoning level indicated that achievement gaps between postformal and concrete, and formal and concrete reasoners at the beginning of the instruction diminished by the end of the instruction. Similarly, other studies also revealed that argumentation-based inquiry instruction helped to close achievement gaps among LAS and HAS (Akkus et al., 2007) and students having a consistent misconception and those having a scientific conception (Acar, 2014). However findings of the previous research did not provide a direct response to whether providing equity to prospective science teachers with different scientific reasoning skills is possible. The result of the present study is promising for ensuring achievement equity among prospective science teachers with different scientific reasoning skills. Nevertheless, the findings also show the situational knowledge gap among reasoners neither closed nor lessened during instruction.

In sum, we found prospective science teachers who are concrete and formal reasoners developed their scientific reasoning and decrease of achievement gaps among prospective science teachers with different reasoning abilities. Former result implies that it is possible to enhance prospective science teachers' not only argumentation skills (Acar, 2008; Zembal-Saul et al., 2002) but also scientific reasoning skills in an argumentation-based inquiry course. On the other hand, contrary to finding of Lewis and Lewis (2008), latter result suggests that it is possible to reduce achievement gaps among students with different reasoning abilities in college.

**Limitations**

There are several limitations in this study. First, although the sample size of each group of reasoners is suitable for doing inferential statistics, to get more compelling results sample sizes would have to be larger. Researchers can use larger sample sizes in future studies to address this limitation. Second, the sample of prospective science teachers in this study may not be representative for the overall population of prospective science teachers since this study took place in one mid-western American university. To test the generalizability of the findings, researchers can carry out a similar study with prospective science teachers in universities which are in different geographic regions. Third, scientific reasoning and conceptual knowledge test used in this study had internal consistencies that were below .70. First of all, internal consistency estimates in this study for scientific reasoning were close to .70 (.69 for pretest and .67 for posttest). In fact several studies also found reliability estimates of this instrument with college students that were below .70 (e.g., Lawson et al., 2000, Schen, 2007). In addition, our results regarding high scientific reasoners' advantage over low scientific reasoners on achievement and situational conceptual knowledge are consistent with the findings of previous research (e.g., Coletta & Phillips, 2005; Johnson & Lawson, 1998; Lawson et al., 2007). This shows that this test gives reliable results in different research contexts. On the other hand, internal consistencies of the two subscales of conceptual knowledge test were .60 and .47. This low reliability of the subscales may threaten the construct validity of the subscales. However, our results regarding significant differences of situational knowledge and no difference of declarative knowledge among reasoners strengthen the construct validity of the subscales because prior research has shown formal reasoners are more skillful in higher order reasoning skills than concrete reasoners (Acar, 2014; Ates & Cataloglu, 2007). In addition to low internal consistency, two conceptual knowledge subscales explained approximately one fourth of the posttest variance. A similar result was also found by Li (2001). More clearly, Li (2001) analyzed science items in Third International Mathematics and Science Study. The author performed logical, factor, and protocol analyses on the data and found that items can be linked to knowledge types (i.e., declarative, procedural, schematic, and strategic knowledge). Although this encouraging result, the author found as ours that two, three and four factor (i.e., knowledge types) solutions of the data explained 21.95, 27.29, and 32.27% of the total variance respectively (pp. 162-166). Nevertheless, since conceptual knowledge test was developed by the authors of this study and not pilot-tested previously, more should be done to improve the internal consistency of the subscales in the conceptual knowledge test. Pilot testing on a larger sample of prospective science teachers can help researchers eliminate the items which do not contribute to either of the conceptual knowledge subscales. Finally, there may be a ceiling effect for the measure of scientific reasoning. Since postformal reasoners started the course with high scientific reasoning scores, it would be unrealistic to expect a significant increase in their scientific reasoning. Thus the result of *t* test analysis for this group is inconclusive from this point of view.

**Implications**

This study shows the promise of an argumentation-based inquiry instruction in reducing the scientific reasoning and achievement gaps among prospective science teachers with different levels of scientific reasoning. Although we expected postformal reasoners would also have developed their scientific reasoning, there may have been a potential ceiling effect for this group. In fact, other high-reasoning students, formal reasoners, developed their scientific reasoning as well as concrete reasoners. Thus we can conclude that this argumentation-based inquiry course was helpful for most of the prospective science teachers in the development of their scientific reasoning. Accordingly, we suggest in accordance with Acar (2014) that argumentation-based inquiry instruction can be utilized in teacher education programs to achieve equity among prospective science teachers. Despite this encouraging result, situational knowledge and achievement gaps still existed at the end of the instruction and did not close completely. First of all, the findings show that students' scientific reasoning level made a difference on their situational knowledge. If we connect this finding with the result of the scientific reasoning gap decrease among reasoners, one might also expect a decline of the gap in situational knowledge, which was not the case. We interpret this to mean that there may be several thresholds of scientific reasoning level which cause differences among groups and these threshold values were not reached by low-level scientific reasoners in the limited time of this one course of argumentation-based inquiry instruction. In summary, we recommend that argumentation and inquiry be incorporated into science curriculum in the early years of education so that it may be more reasonable to expect closure of scientific reasoning, situational knowledge, and achievement gaps among prospective science teachers by this prolonged engagement.

## References

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N.G., Mamlok-Naaman, R., Hofstein, A.,…& Tuan, H.L. (2004). Inquiry in science education: International perspectives. *Science Education, 88*(3), 397-419. http://dx.doi.org/10.1002/sce.10118

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. (2014). Scientific reasoning, conceptual knowledge, & achievement differences between prospective science teachers having a consistent misconception and those having a scientific conception in an argumentation-based guided inquiry course. *Learning and Individual Differences, 30*, 148-154. http://dx.doi.org/10.1016/j.lindif.2013.12.002

Acar, O., Turkmen, L., & Roychoudhury, A. (2010). Student difficulties in socio-scientific argumentation and decision-making research findings: Crossing the borders of two research lines. *International Journal of Science Education, 32*(9), 1191-1206. http://dx.doi.org/10.1080/09500690902991805

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. http://dx.doi.org/10.1080/09500690601075629

Ates, S., & Cataloglu, E. (2007). The effects of students' reasoning abilities on conceptual understandings and problem-solving skills in introductory mechanics. *European Journal of Physics, 28*(6), 1161-1171. http://dx.doi.org/10.1088/0143-0807/28/6/013

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. http://dx.doi.org/10.1080/095006900412284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Psychology Press.

Coletta, V.P., & Phillips, J.A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics, 73*(12), 1172-1182. http://dx.doi.org/10.1119/1.2117109

de Jong, T., & Ferguson-Hessler, M.G.M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105-113. http://dx.doi.org/10.1207/s15326985ep3102_2
Dogru-Atay, P., & Tekkaya, C. (2008). Promoting students' learning in genetics with the learning cycle. *The Journal of Experimental Education*, *76*(3), 259-280. http://dx.doi.org/10.3200/JEXE.76.3.259-280
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. http://dx.doi.org/10.1002/tea.20248
Gerber, B.L., Cavallo, A.M.L., & Marek, E.A. (2001). Relationships among informal learning environments, teaching procedures and scientific reasoning ability. *International Journal of Science Education*, *23*(5), 535-549. http://dx.doi.org/10.1080/09500690116971
Giere, R.N. (1984). *Understanding scientific reasoning*. New York, NY: College Publishing.
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. http://dx.doi.org/10.1080/09500690110049150
Jackson, J.K., & Ash, G. (2012). Science achievement for all: Improving science performance and closing achievement gaps. *Journal of Science Teacher Education*, *23*(7), 723-744. http://dx.doi.org/10.1007/s10972-011-9238-z
Jensen, J.L., & Lawson, A. (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE-Life Sciences Education*, *10*(1), 64-73. http://dx.doi.org/10.1187/cbe.10-07-0089
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*(6), 757-792. http://dx.doi.org/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*(3), 287-306. http://dx.doi.org/10.1007/s10972-009-9134-y
Johnson, M.A., & Lawson, A.E. (1998). What are the effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes. *Journal of Research in Science Teaching*, *35*(1), 89-103. http://dx.doi.org/10.1002/(SICI)1098-2736(199801)35:1<89::AID-TEA6>3.0.CO;2-J
Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, *14*(2), 169-175. http://dx.doi.org/10.1002/tea.3660140212
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. http://dx.doi.org/10.1080/0950069980200707
Klaczynski, P.A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, *71*(5), 1347-1366. http://dx.doi.org/10.1111/1467-8624.00232
Kuhn, D. (1991). *The skills of argument*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571350
Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810-824. http://dx.doi.org/10.1002/sce.20395
Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.

Kuhn, D., & Dean, D., Jr. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development, 52*(2), 261-288. ttp://dx.doi.org/10.1207/s15327647jcd0502_5

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*(4), 435-451. http://dx.doi.org/10.1016/j.cogdev.2008.09.006

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

Kuhn, T.S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: The University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226458106.001.0001

Lawson, A.E. (1978). The development and validation of a classroom test of scientific reasoning. *Journal of Research in Science Teaching*, *15*(1), 11-24. http://dx.doi.org/10.1002/tea.3660150103

Lawson, A.E. (2003). The nature and development of hypothetico-predictive argumentation with implications for science education. *International Journal of Science Education, 25*(11), 1387-1408. http://dx.doi.org/10.1080/09500690320000052117

Lawson, A.E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry. *Journal of Research in Science Teaching, 42*(6), 716-740. http://dx.doi.org/10.1002/tea.20067

Lawson, A.E. (2010). Basic inferences of scientific reasoning, argumentation, and discovery. *Science Education, 94*(2), 336-364.

Lawson, A.E., Banks, D.L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching, 44*(5), 706-724. http://dx.doi.org/10.1002/tea.20172

Lawson, A.E., Clark, B., Cramer-Meldrum, E., Falconer, K.A., Sequist, J.M., & Kwon, K.J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching, 37*(1), 81-101. http://dx.doi.org/10.1002/(SICI)1098-2736(200001)37:1<81::AID-TEA6>3.0.CO;2-I

Lawson, A.E., & Weser, J. (1990). The rejection of nonscientific beliefs about life: Effects of instruction and reasoning skills. *Journal of Research in Science Teaching, 27*, 589-606. http://dx.doi.org/10.1002/tea.3660270608

Lawson, A.E., & Worsnop, W.A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. *Journal of Research in Science Teaching, 29*(2), 143-166. http://dx.doi.org/10.1002/tea.3660290205

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. http://dx.doi.org/10.1002/tea.20254

Li, M. (2001). A framework for science achievement and its link to test items. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Liao, Y.W., & She, H.C. (2009). Enhancing eight grade students' scientific conceptual change and scientific reasoning through a web-based learning program. *Educational Technology & Society*, *12*(4), 228-240.

Marušić, M., & Sliško, J. (2012). Influence of three different methods of teaching physics on the gain in students' development of reasoning. *International Journal of Science Education, 34*(2), 301-326. http://dx.doi.org/10.1080/09500693.2011.582522

McDermott, L.C. (1996). *Physics by inquiry: An introduction to physics and the physical sciences***.** New York, NY: Wiley.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
NGSS Lead States. (2013). *Next generation science standards: For States, by states*. Washington, DC: The National Academies Press.
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994-1020. http://dx.doi.org/10.1002/tea.20035
Root-Bernstein, R.S. (1989). How scientists really think. *Perspectives in Biology and Medicine, 32*(4), 472-488. http://dx.doi.org/10.1353/pbm.1989.0017
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*(4), 387-409. http://dx.doi.org/10.1080/0950069032000119456
Schen, M.S. (2007). *Scientific reasoning skills development in the introductory biology courses for undergraduates.* Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.
She, H.C., & Liao, Y.W. (2010). Bridging scientific reasoning and conceptual change through adaptive web-based learning. *Journal of Research in Science Teaching, 47*(1), 91-119. http://dx.doi.org/10.1002/tea.20309
The Organisation for Economic Co-operation and Development. (2013). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed (Volume II)*. Paris: OECD publishing.
Watson, J.R., Swain, J.R.L., & McRobbie, C. (2004). Students' discussions in practical scientific inquires. *International Journal of Science Education, 23*(1), 25-45. http://dx.doi.org/10.1080/0950069032000072764
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.
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*(3), 343-367. http://dx.doi.org/10.1002/sce.10025
Zembal-Saul, C. (2009). Learning to teach elementary school science as argument. *Science Education, 93*(4), 687-719. http://dx.doi.org/10.1002/sce.20325
Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education, 32*(4), 437-463. http://dx.doi.org/10.1023/A:1022411822951
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. http://dx.doi.org/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. http://dx.doi.org/10.1002/tea.10008