Volume 25, Number 2 • Summer 2021

ISSN: 1087-3430

Electronic Journal for Research in Science & Mathematics Education

Flagship Journal of the International Consortium for Research in Science & Mathematics Education (ICRSME)

ICRSME

PUBLISHER
ICRSME, International Consortium for Research in Science & Mathematics Education

Cover designed by Dusty Crocker

CONTENTS

Volume 25 No. 2 | Summer 2021

ICRSME Sponsor

TCU®

COLLEGE OF
EDUCATION

ANDREWS INSTITUTE

Support Us

ICRSME is a nonprofit organization and your donations will help fund its various activities including the publication of EJRSME and hosting international consultations and virtual meetings.

https://icrsme.com/support-us

ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 25, NO. 2, 1-5

ICRSME
International Consortium for Research in Science & Mathematics Education

Editorial: The Characteristics of a Good Listener

Mark A. Bloom
Dallas Baptist University

Sarah Quebec Fuentes
Texas Christian University

In a previous editorial, we addressed public mistrust of science and mathematics (Bloom & Quebec Fuentes, 2020a). We argued that this mistrust stems from various sources, including political partisanship, religious conflict, and historical marginalization of various groups as well as a lack of confidence in experts. One way to enhance trust in science and mathematics is for experts to purposefully consider how they convey information to the public (Oreskes, 2014). Oreskes further makes the point that simultaneously everyone (scientists, mathematicians, educators, and lay people alike) needs to be a good listener. We closed the editorial with a mandate for science and mathematics educators to contribute to the efforts in reestablishing trust in science and mathematics by developing good listeners. The rest of the present editorial delves into the characteristics of a good listener.

A good listener needs to have an understanding of the natures of science and mathematics. To trust information conveyed by scientists and mathematicians, the general public needs to have an understanding of the characteristics of science, the scientific process, and mathematical modeling (Bloom & Quebec Fuentes, 2020b). Understanding the tentative, subjective, and communal nature of science as well as the mathematical modeling process provides a foundation for understanding information conveyed by scientists and mathematicians.

A prime example of this need is the public criticism of Dr. Anthony Fauci for his changing recommendations on mask-wearing over the course of the COVID-19 pandemic. Early in the pandemic, when hospitals were experiencing shortages of personal protective equipment (PPE) and when there was little data regarding disease transmission, Fauci advised the public that masks were not necessary. Later, when new evidence revealed that asymptomatic carriers of COVID-19 could indeed spread the disease and in light of a tremendous increase in PPE production, his recommendation changed, advocating for universal mask-wearing to help control disease spread (Sonnemaker, 2021). In an interview with Kara Swisher (2021) on the *New York Times* "Sway" podcast, Fauci defended his evolving recommendations regarding masks indicating that the people who are criticizing him as a 'flip flopper' who is misleading the public actually lack an informed understanding of science; to them he says, "let me give you a flash. That's the way science works. You work with the data you have at the time" (Allen, 2021). He further emphasized that in science, one must be "humble enough and flexible enough to change with the data." If the public had a better understanding of how scientific recommendations continue to change with new data, perhaps they could recognize that the changes reflect a developing understanding of the scientific phenomenon - and perhaps they could *better listen* and learn.

A good listener needs to identify who is an expert. While Dr. Fauci is certainly an expert, one does not have to hold a Ph.D. to be an expert. Each of us may have developed expertise in a particular field, but few, if any, possess expertise in all areas. Nichols (2017) indicates that through metacognition, people are able to evaluate their abilities and identify those areas that they are quite good at and those which require help from others. While I (Mark) have a fairly developed understanding of North Texas wildlife and can identify much of the local flora and fauna, at least to a family level, I still rely upon field guides and other resources to be sure of my identifications - I am simply not an expert. Far less do I understand what goes on under my car's hood, how my air conditioner properly cools my home, or how the Wifi communicates with all of my family's electronic devices. When these fail to operate properly, I seek the advice of an auto mechanic, an HVAC technician, or an electrician. Nichols (2017) asserts that many "unskilled or incompetent people overestimate their abilities far more than others" because they do not have that critical metacognitive skill to recognize their lack of expertise; thus, they may never seek expert opinions (p. 45). Further, when expert advice does not reflect what they want to hear, Nichols maintains that bad listeners "search for the loopholes in expert knowledge that will allow them to disregard all expert advice they don't like" (p. 23). In contrast, good listeners recognize when they need the help of experts and carefully consider the advice they offer.

A good listener needs to determine whether information conveyed is from credible sources. Over the last 30 years, the number of sources for news and information has increased drastically, paralleled by an increase in political partisanship of sources and false information (Iyengar & Massey, 2019). Even good listeners who recognize their limitations and are willing to listen to the advice of experts can end up with bad information. The Stanford History Education Group (SHEG; 2016) conducted a study of students' civic online reasoning to determine how well young people evaluate the information they access on their smartphones and computers. After analyzing 7,804 student responses, SHEG determined that the vast majority were easily duped into accepting misinformation as fact. In one of the tests, involving high school and college students, the task was to discern that a website they visited presenting information on minimum wage policy was for an organization that was a front group for a D.C. lobbyist and, therefore, presented partisan opinion - only nine percent of the high schoolers and only seven percent of the college students made this discovery (SHEG, 2016). Make no mistake; adults also fall prey to well-disguised misinformation (Gottfried & Grieco, 2018).

The SHEG researchers identified three distinct strategies that separated the discriminating students from the rest - strategies that made them *good listeners* (Wineburg & McGrew, 2016). First, when *good listeners* land on an unfamiliar website, they open a new browser and Google the name of the sponsoring organization or its leaders - an approach the authors refer to as *lateral reading.* Second, *good listeners* do not rely upon the 'About' page on websites, understanding that they cannot determine the validity of a website based upon its own description of itself. Third, *good listeners* scroll through all search results before determining where to click first. Those who are less discerning often assume the order of the results somehow equated to their reliability. With information available in such large quantities and from so many sources, the populace must be critical consumers of information, now more than ever.

A good listener needs to demonstrate healthy skepticism and demand additional information when it is not provided. When I (Sarah) was younger, I overheard my relatives talking about their challenges with the symptoms of menopause and the risks of hormone replacement therapy (HRT). Based on this conversation, I believed that HRT was not a viable option for women. However, the story is more complex. In 2002, the results of a study on HRT indicated that hormonal treatment for menopausal symptoms increased risk of cardiovascular disease and breast cancer. Later studies suggested that the resulting decline in HRT prescription and use (Cagnacci & Venier, 2019) over the subsequent years could have contributed to tens of thousands of premature deaths among women (Sifferlin, 2013). A good listener would have probed a bit deeper before abandoning their medications. While true that the treatment group (those receiving HRT) did show an increased risk of cardiovascular disease and breast cancer, the actual numbers tell a less grim story. Compared to the control group, among the 10,000 women in the treatment group, there were only eight additional strokes, seven additional cardiac events, and eight additional cases of breast cancer (Prescrire Int., 2003). A good listener might have asked questions that would inform her if she was at risk of being one of those additional cases - questions like 'Do I have a family history of cardiovascular disease or breast cancer?' or 'Do I have any preexisting conditions that increase my susceptibility to cardiovascular disease or breast cancer?' Such questions help good listeners make the most appropriate choices.

A good listener needs to consider how their lived experiences may influence how they interpret information. Iyengar and Massey (2019) argue that the aforementioned changes with news outlets and the political divide overpower scientists and mathematicians attempts at communicating effectively. In contrast, Jamil Zaki, a professor of psychology at Stanford University, explains that empathy is a skill that can be developed to address the "intergroup empathy gap" (Santos, 2020). One strategy for enhancing empathy is to "disagree better" and "cultivate curiosity" by communicating and listening to each other's stories about the origins of one's beliefs (Zaki, n.d.).

An example of such is the ongoing dialogue between Dr. Deb Haarsma, President of BioLogos, and Dr. Hugh Ross, President of Reasons to Believe. According to its website, BioLogos "invites the church and the world to see the harmony between science and biblical faith." The Reasons to Believe website informs that the mission of the organization is to "make every effort to help people discover that sound reason and scientific research consistently affirm the truth of the Bible and of the Good News it reveals." Both leaders are evangelical Christian and both are trained in astrophysics, yet the organizations they lead hold quite distinct views on the origin of species on earth and on how scripture should be interpreted. In *Discussing Origins with Reasons to Believe and BioLogos* (Keathley et al., 2017), each leader describes how their organizations differ. Haarsma describes a key difference as the two groups' approach to biblical inerrancy. According to Haarsma the "tent" of BioLogos,

> includes a range of views on inerrancy. Some actively embrace the term, viewing the Bible as inerrant in matters of faith and practice. Others, while taking Scripture seriously as authoritative and inspired, do not find *inerrant* to be a helpful term in describing their views. (Keathley et al., 2017, p. 13)

By contrast, Ross, emphasizes that Reasons to Believe holds a "strong commitment to biblical inerrancy," and

> den[ies] that Scripture should be required to fit alien preunderstandings inconsistent with itself, such as naturalism, evolutionism, scientism, secular humanism, and relativism. (Keathley et al., 2017, p. 15)

With these two distinct a priori assumptions about biblical inerrancy, each views new scientific claims from completely different perspectives. BioLogos might see new discoveries as informative in adjusting how scripture should be interpreted, while Reasons to Believe will use biblical interpretation to deem the science as credible or not. Despite having such opposing views to how one interprets the same biblical text that they both hold as vital to their faith, Haarsma and Ross, because they are good listeners, maintain dialogue and seek to understand each other. Understanding why each other holds the beliefs they do, and understanding how these beliefs are an outcome of their different lived experiences, allows them to, as Haarsma describes, "celebrate [their] common commitment to biblical Christianity and to science as a means of understanding God's creation" (Keathley et al., 2017, p. 12) rather than focus on what divides them.

Different lived experiences, especially those rooted in religious beliefs such as those of Haarsma and Ross, can have great influence on how individuals interpret, understand, and accept science and mathematics. This last example provides an opportunity to share information about our next issue of *EJRSME*, a special issue with contributions from Sinai and Synapses fellows. The mission of Sinai and Synapses is to "offer people a worldview that is both scientifically grounded and spiritually uplifting" and to "provide tools and language for learning and living to those who see science as their ally as they pursue personal growth and the repair of our world." The special issue will highlight the work of some of these fellows as they communicate science and mathematics in formal and informal settings to both science and religious audiences. Topics will include climate change education, environmental racism, Judaism's embrace of science, terror management theory, racial equity in science and mathematics, health education during the COVID-19 pandemic, and science and religion as distinct ways of knowing. This special issue will continue the ongoing conversation about effective communication and listening in science and mathematics.

References

Allen, M. (2021, June 20). Scoop: Fauci's offensive against "craziness." *Axios.* https://www.axios.com/faucis-offensive-craziness-0dbcc643-140a-4407-96d9-eceb92a1289c.html

BioLogos. (n.d.). *About us.* https://biologos.org/about-us#our-mission

Bloom, M. A., & Quebec Fuentes, S. (2020a). Addressing public mistrust of science and mathematics. *Electronic Journal for Research in Science & Mathematics Education, 24*(4), 1-4.

Bloom, M. A., & Quebec Fuentes, S. (2020b). How the COVID-19 pandemic reveals gaps in science and mathematics instruction. *Electronic Journal for Research in Science & Mathematics Education, 24*(2), 1-5.

Cagnacci, A. & Venier, M. (2019, September). The controversial history of hormone replacement therapy. *Medicina.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6780820/

Gottfried, J., & Grieco, E. (2018, October). Younger Americans are better than older Americans at telling factual news statements from opinions. *Pew Research Panel.* https://www.pewresearch.org/fact-tank/2018/10/23/younger-americans-are-better-than-older-americans-at-telling-factual-news-statements-from-opinions/

Iyengar, S. & Massey, S. M. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences, 116*(16), 7656-7661.

Keathley, K., Stump, J. B., & Aguirre, J. (2017). *Discussing origins with Reasons to Believe and BioLogos.* InterVarsity Press.

Nichols, T. (2017). *The death of expertise: The campaign against established knowledge and why it matters.* Oxford University Press.

Oreskes, N. (2014, May). Why should we trust scientists? [Video]. TED Conferences. https://www.ted.com/talks/naomi_oreskes_why_we_should_trust_scientists#t-990833

Prescrire Int. (2003, April). Postmenopausal hormone therapy: cardiovascular risks. (2003). *Prescrire international*, *12*(64), 65–69. https://pubmed.ncbi.nlm.nih.gov/12674130/

Reasons to Believe. (n.d.). *About Reasons to Believe.* https://reasons.org/about

Santos, L. (Host). (2020, June 21). The war for kindness [Audio podcast episode]. In *The Happiness Lab.* Pushkin Industries. https://www.happinesslab.fm/season-2-episodes/episode-9-the-war-for-kindness

Sifferlin, A. (2013, July 20). Hormone-replacement therapy: Could estrogen have saved 50,000 lives? *Time.* https://healthland.time.com/2013/07/20/hormone-replacement-therapy-could-estrogen-have-saved-50000-lives/

Sinai and Synapses. (n.d.). *Our mission and methods.* https://sinaiandsynapses.org/our-mission-and-methods/

Sonnemaker, T. (2021, June 20). Fauci pushes back against the 'craziness' of attacks on his evolving pandemic response, saying 'that's the way science works.' *Insider*. https://www.businessinsider.com/fauci-defends-himself-against-craziness-of-attacks-masks-pandemic-axios-2021-6

Stanford History Education Group (2016, November 22). *Evaluating information: The cornerstone of civic online reasoning. Executive summary*. Stanford History Education Group.

Swisher, K. (Host). (2021, June 21). Anthony Fauci on the lab leak theory and emailing Mark Zuckerberg [Audio podcast episode]. In *Sway*. The New York Times. https://www.nytimes.com/column/sway

Wineburg, S. & McGrew, S. (2016, November 1). Why students can't Google their way to truth. *EdWeek*. https://www.edweek.org/teaching-learning/opinion-why-students-cant-google-their-way-to-the-truth/2016/11

Zaki, J. (n.d.) *Challenge 3 - Disagree better*. The War for Kindness. Retrieved June 29, 2021, from https://www.warforkindness.com/challenge-3-disagree-better

ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 25, NO. 2, 6-29

Investigating Weather, Climate, and Climate Change Understanding of Appalachian Middle-Level Students

Tina J. Cartwright
Marshall University

Deb Hemler
Fairmont State University

Paula A. Magee
Indiana University—Purdue University Indianapolis

ABSTRACT

Climate change is an increasingly pervasive global topic, but how much of this discussion is accurately understood by students? Fully comprehending the small fluctuations associated with long term changes in temperature and precipitation is a daunting task for the general public, let alone for middle-level adolescents. This study examines students' understanding of weather, climate and climate change. Forty-seven students, ages 12-14 from the Appalachian region of the US, were surveyed before, immediately after, and six months after a standards-based unit of instruction. The study utilized a questionnaire developed by Boon (2009) with additional questions related to weather and climate. Qualitative data were analyzed using a constructivist framework and student responses were examined for understanding of the main content ideas. The students' understandings were analyzed over time for shifts and were also compared with previously published research (Bodzin et al., 2014; Boon, 2009). Students made improvements in some aspects of understanding with instruction but not all gains persisted to six months post instruction. Students' distinctions between weather and climate were altered by instruction, persisted, and continued to improve with time. Students demonstrated a general understanding of the differences between weather and climate but struggled when asked to apply this knowledge to specific situations. Some improvements in students' basic understanding of the greenhouse effect were evident, but some of these improvements degraded with time. While instruction was able to temporarily improve understanding of greenhouse gases, and the benefits of the greenhouse effect, overall students did not retain this understanding over the long term.

Keywords: earth science education, longitudinal study, scientific literacy, climate literacy

Introduction

One of the most important challenges facing the citizens of the 21st century will undoubtedly be climate change. Yet, understanding climate change remains problematic (Dawson, 2015; Khalid, 2003). Distinguishing small fluctuations from long term changes in temperature and precipitation is challenging for the general public, let alone for the adolescent (Harker-Schuch & Bugge-Hendrikson, 2013; Lambert et al., 2012). Unfortunately, students may only receive instruction on this general environmental science topic in middle-school and in a general science course during their freshman year of high school. This study examines the understanding of weather, climate and climate change, of Appalachian middle-school students, before, immediately after, and 6 months after a standards-based unit of instruction prior to adoption of the Next Generation Science Standards (NGSS). Future research will monitor how these middle-school students' understanding 1) evolves through high school and 2) compares with students in this area after adoption of NGSS.

A better understanding of what students know during the adolescent age, when they receive the most instruction on environmental issues, can be helpful for scientists and educators as they prepare scientifically literate adults. To this point, two studies (Boon, 2009; Bodzin, Anastasio, Sahagian, Peffer, Dempsey, & Steelman, 2014) provided useful research in this area. However, both studies assessed student understanding at one point in time only. The current study adds to the research base because it assesses students' longitudinal knowledge (*before*, *immediately after, and 6 months after* a standards-based unit of instruction) of the earth-atmosphere system. Additionally, this research study positions the analysis within a constructivist framework, where students are recognized as being owners of their own knowledge (Carr, Barker, Bell, Biddulph, Jones, Kirkwood, Pearson and Symington, 2013; Von Glasersfeld, 2013). This positioning allows for an analysis of student thinking and subsequent recommendations for classroom teachers regarding productive instructional practices. As such, this research provides preliminary data analyses that support continued research to further investigate students' understanding of weather, climate and climate change as they progress through high school.

Background

Student Distinctions Between Weather and Climate

Weather describes the atmospheric conditions over short-term duration (minutes, hours, days, months, and years), while climate describes these same conditions averaged over at least a 30-year period over a much wider area (National Climatic Data Center, 2008). Researchers have established that both middle-school and high-school students have problems distinguishing between weather and climate (Papadimitriou, 2004; Lombardi & Sinatra, 2012). Dawson (2015) found that high school students use the terms 'weather' and 'climate' interchangeably. Bodzin et al. (2014) found that the majority of middle-school students recognized that climate changes at a much slower rate than weather, but few students understand that a region's climate describes a region's weather conditions and that this changes on the order of decades. The majority of middle-school students felt that "climate is defined as weather patterns that change on a scale of at least a few *weeks*" rather than the correct response of *decades* (Bodzin, et al., 2014). Undoubtedly, an understanding of the greenhouse effect and climate change requires that students be able to differentiate between weather and climate (Jarrett, Ferry, & Takacs, 2012).

It is only natural that people use their experiences with local weather to make critical inferences about global climate (Read et al., 1994). Read et al. (1994) suggested that people do, in fact, use their short-term weather experiences (like heat waves or cold spells) to make judgments about longer term climate trends. Similarly, preservice elementary teachers cited recent weather events (such as a summer heat wave) to serve as evidence of global warming (Papdimitriou, 2004). Other research has shown that 60% of high school students indicated that "climate often changes from year to year" (Gowda et al., 1997, p. 2236). Additionally, 15% of these same high school students indicated that they had personally witnessed evidence of climate change. Gowda et al. (1997) claimed that these evidences of climate change were "memorable weather events" (p. 2236) (e.g. a flooding event, a hot summer, or lack of snow at Christmas).

Additional confusion between weather and climate may result from a students' perception and understanding of deep time (Lombardi & Sinatra, 2012). Deep time is often referred to as geological time and has been shown to be challenging for students to understand (Libarkin et al., 2005; Prather, 2005). Lombardi and Sinatra (2012) examined undergraduate students to determine if there was a relationship between their understanding of deep time and their distinctions between weather and climate. They found that a greater knowledge of deep time and improved perceptions of human-induced climate change explained a significant portion of the variance in students' understanding of weather and climate distinctions. Nevertheless, students have been shown to improve their differentiation between weather and climate with a relatively brief intervention (Lombardi & Sinatra, 2012).

Student Understanding of Climate Change Issues

Middle and high school students often confuse climate change with unrelated environmental issues and therefore, have a limited understanding of environmental responses to climate change (Bofferding & Kloser, 2015). Students understand that carbon dioxide plays an important role in climate change but are not familiar with other greenhouse gases (GHG). Specific alternative conceptions include naive understanding about increases in GHG and ozone depletion (Bodzin et al., 2014; Bostrom et al., 1994; Rye, Rubba, and Wiesenmayer, 1997), inability to identify GHG (Bodzin et al., 2014;Bofferding & Kloser, 2015), GHG distribution in the atmosphere (Bodzin & Fu, 2013), and global climate change impacts on other Earth systems (Shepardson et al., 2009). Additionally, students often attribute air pollution or acid rain to climate change (Bofferding & Kloser, 2015). Students lack the complex understanding of climate change and often demonstrate an oversimplified understanding as a "unidirectional linear cause-effect" model (Shepardson et al., 2014). Additionally, students struggle with identifying and associating appropriate actions that might reduce climate change (Bodzin et al., 2014, Boyes & Stanisstreet, 1993; Boyes et al., 2009; Kilinc et al., 2011). These studies have shown some success with actions such as reducing car usage, using more fuel-efficient cars, less electricity and more alternative energy sources but remaining challenges associated with litter, pollution, endangered species, insecticides and nuclear energy.

These issues in student understanding have been confirmed in many international studies throughout many different countries. Dawson's study (2015) revealed that Australian students identified carbon dioxide as the only GHG. Likewise, Fisher (1998) reported that Australian students associated GHG with the ozone layer. Similarly, Boon (2009) compared Australian students with British students and that both held similar misconceptions related to GHG and climate change. Additionally, Australian, Norwegian and Turkish students held some level of understanding of the greenhouse effect but held misconceptions regarding both the ozone layer and greenhouse effect (Kilinc et al., 2011).

Impact of Instruction on Climate Change Understanding

Even though students' understanding about climate change has been well documented, research literature suggests that these misconceptions can be modified through effective instruction (Bodzin & Fu, 2013). Lectures on climate change were shown to slightly improve Austrian and Danish students' understanding about climate change (Harker-Schuck & Bugge-Hendrikson, 2013). Visualizations and virtual experiments were shown to be effective in improving Year 6 student understanding of global climate change (Varma & Linn, 2011). A three-week intervention using a variety of instructional techniques that focused on climate change being a socioscientific issue statistically improved Year 10 student understanding (Klosterman & Sadler, 2010). Similarly, McNeill and Vaughn (2012) found similar results in Year 11/12 students after an 11-lesson unit which targeted climate change and environmental action. Middle-school climate change instruction which utilized critical evaluation and plausibility appraisal promoted greater understanding of socio-scientific topics and increased use of scientific thinking when considering alternative explanations (Lombardi, Brandt, Bickel, & Burg, 2016). Similar to this current study, all of these studies relied on heavy collaboration between the researcher and teacher to deliver effective instruction. While many studies show changes in student understanding of climate change, measures of the persistence of that knowledge months after instruction, is noticeably absent in the literature.

Research Focus and Questions

This study compared students' understanding of weather, climate and climate change before (t_1) a unit of instruction on these topics, immediately after (t_2), and six months post instruction (t_3) which took place in 2011-2012 before the NGSS were adopted in 2016.
This study explored the following two research questions:

(1) What are Appalachian middle-level students' understandings about weather, climate and climate change?

(2) Following a standards-based unit of instruction, how do these understandings persist over time: immediately post instruction (t_2) and after 6 months of time (t_3)?

Methodology

Participants

Forty-seven students between the ages of 12-14 years old participated in this study. The middle school is located in a suburban area of a mostly rural Appalachian state with a total enrollment of approximately 600 students of which 95% identify as white. As it is imperative to establish the context and influence of community, this middle school is located within a state that is predominantly supported by the coal industry. Obtaining approval for curriculum standards addressing climate change has proven to be a challenge given the influence of extractive industries and the dependence of residents on them for their livelihoods. This context develops unique and important cultural experiences for students in this region and naturally affects their perceptions about the consequences of mining, extracting and processing coal. The researchers recognize the potential impacts here on students' understanding about topics such as climate change and this is discussed more in the results section. The targeted student population included a range of academic abilities but did not include any students with identified disabilities because the school did not have access to sufficient support services for all academic teams. Students were placed in these academic teams based upon two factors: 1) the lack of an identified learning disability and 2) their math course requirements. Three teams were included in this study which included one team of 7th graders taking Algebra I and two teams of 8th graders. The majority of the 8th graders were taking Algebra I, but a few were enrolled in Geometry from the nearby high school.

Assessment Measure Development

Previously published research provided the international comparison dataset as well as the majority of the assessment measure. The Boon study (2009) included the following: 168 year 10 (ages 14-15) and 183 year 8 (ages 12-13) students from a northern UK city and 79 year 8 (also ages 12-13) students and 310 year 10 (ages 14-15) students from four schools in a Queensland, Australia city. Both of these studies in Australia and the UK were conducted when there was a high level of media coverage of the phenomena in each of the countries due to unseasonal weather patterns and political debates taking place in both countries. Relatively little media coverage was present in the US as the political climate was focused on economic issues associated with the recession of 2011. Boon developed the initial assessment measure after two pilot tests were conducted on a class of year 8 students in the UK (2009). Boon examined the student responses and selected the most suitable questions from the two trial administrations.

The present study utilized Boon's previously published questions (Boon, 2009) which included multiple choice, yes/no, and constructed response type questions. To add depth to the assessment measure, additional questions related to weather and climate, not included in Boon's study, were developed and included. These additional weather and climate questions were developed by the lead author who is a meteorologist, and the questions were read for face validity (Creswell, 2008) by meteorologists at the National Weather Service. The questions push the participant to think beyond memorized definitions and were developed to ascertain participant understanding of real-world application differences between weather and climate. The complete assessment measure used in this study can be found in Appendix A.

The assessment measure was given to 47 year 7 and 8 students (ages 12-14) at three different time periods during the 2011-2012 academic year: 1) t_1 - before instruction, 2) t_2 - immediately post instruction, and 3) t_3 - six months after instruction (delayed post) by the students' usual science teacher. The portion of students with parental consent also participated in semi-structured group interviews at the delayed post time period (t_3). The interview questions can be found in the Appendix B. The group interviews were used to add clarity and depth to the student responses on the weather and climate questions particularly. Students were a convenience sample based on the willingness of the classroom teacher to provide instructional time for the lead author to provide the unit of instruction (described below).

A total of 20 questions were used on the assessment measure, some of which, were divided into multiple prompts resulting in 43 items requiring student responses. Of these items, 22 were used in this study. To tease out student understanding of aspects of weather and climate change, these questions were grouped into three categories: weather and climate (9 items), human actions and the greenhouse effect (6 items), and greenhouse gases (7 items). Paired-comparison two-tailed t-tests were performed between each administration of the instrument with the basic assumption that the differences in scores were normally distributed in a class with little to no instruction on climate change. A confidence interval of 95% was chosen to determine the mean differences. The tests were used to compare student knowledge of climate change prior to instruction (t_1 - pretest) with knowledge post instruction (t_2 - posttest), student knowledge post instruction (t_2- posttest) with knowledge six months post instruction (t_3- delayed posttest), and finally student knowledge prior to instruction (t_1- pretest) with knowledge six months post instruction (-t_3 delayed posttest). Normalized gains ($<g>$) and effect sizes were calculated overall and for each of the assessment measure categories: weather vs climate and greenhouse effect to indicate effectiveness of instruction in promoting conceptual understanding. The "average of gains" method was used since it was possible to match the student data.

Interviews

After completing the 6-months post instruction assessment measure (t_3), students participated in small group semi-structured interviews with the same single member of the research team. The semi-structured group interviews lasted 30 minutes and questions to the participants focused on the weather and climate statements particularly. Four sets of interviews with four participants each were conducted for a total of 16 students (those students who had submitted a parental permission form for the interviews). All interviews were audio recorded and transcribed. The primary focus and purpose of the interviews was two-fold: 1) to gather feedback on the weather and climate statements which were developed for this study and 2) provide opportunity for students to vocalize their thinking about these topics which would help us interpret their survey results. The semi-structured group interview protocol is provided in Appendix B.

Themes found in responses to the interview questions provided insights into questionnaire responses. As described by Cohen, Marion and Morrison (2002), a content analysis was performed on the interview transcripts. As is customary with content analysis, “categories are usually derived from theoretical constructs or areas of interest devised in advance of the analysis” (p.475). The initial categories used for this study were: greenhouse effect, greenhouse gases, climate, climate change and weather. Members of the research team coded the transcripts independently looking for both “correct response” as well as student statements that revealed the reasoning connected to the response for the predetermined categories. Developing and using these primary categories allowed the researchers to focus in on the most relevant remarks made in the interviews. Upon completion of the content analysis, researchers re-examined the data pieces selected and looked for overlap and resonance with regard to major themes and understandings. For example, one theme that emerged was the idea that what happens in nature is cyclical and “just happens.” This idea was expressed by several students and often cited as a reason for natural phenomena. These data were then used to better understand some of the quantitative responses linked to understanding of climate change. Appendix C contains a diagram of the coding scheme with sample quotes included for further clarity.

Unit of Instruction

In the fall of 2011, a 10-day unit of instruction was provided by the lead author who served as a temporary teacher to the 12- to 14-year-old students. The unit was based on the state standards and NSES for grades 5-8 which were in effect at the time. The unit of instruction was based primarily on the instructional standards for the state rather than on the assessment measure which had been developed by Boon. To avoid "teaching to the test," the unit was developed with the most relevant concepts as described in the state and national standards. As expected some of these concepts are not included on the assessment measure, but still do play a role in supporting students' content knowledge development in related areas. Likewise, some aspects of the assessment measure were not directly addressed because they were not prevalent in the state standards. This paper will focus on those aspects of the assessment measure which were included in instruction.

At the time of the research, climate change and the greenhouse effect were absent from both state and national standards in the middle school curriculum. Some additional instruction was provided to students which went beyond the expected middle school curriculum regarding the greenhouse effect particularly. The Appendix D presents a comparison of the state standards, the lesson’s essential question, the student learning objectives, a summary of the lesson, and the corresponding question number on our assessment measure. Each lesson was presented over a 2-day period. The majority of instruction (80% instructional time) focused on those concepts directly related to weather and climate which were clearly specified in both state and national standards, which did not necessarily have questions addressing these topics in Boon’s original assessment measure. The remaining instruction targeted ideas related to GHG and climate change, particularly the role that GHG have in mitigating day vs. nighttime temperatures and how they may impact climate change which extended beyond the minimum state standards. Data and evidence of changing amounts of carbon dioxide were also shown and discussed to launch student thinking regarding our atmosphere with an enhanced greenhouse effect.

The five-lesson unit of instruction was based on the 5E Learning Cycle (Bybee et al., 2006). Students were initially engaged with the activation of prior knowledge (Engage phase), then they actively collected evidence as they explored (Explore phase). In the third E students made sense of their evidence by building new scientific explanations (Explain phase) and then they were given a new situation where they applied their new understanding (Elaborate phase). Finally, students’ understanding was evaluated throughout the lesson to determine if further instruction was necessary (Evaluate phase).

Findings and Results

As described previously, our assessment measure naturally divides into three groups of themed questions which focus on: differentiating weather and climate, the greenhouse effect, and greenhouse gases. The following discussion subheadings are based on these themes.

Weather vs. Climate

Nine questionnaire statements required students to choose either weather or climate as the cause of various phenomena. The percentage and frequency (n) of correct responses for each statement is provided in Table 1. On the pretest, students correctly identified weather or climate prompts 63% of the time. Students demonstrated prior knowledge on only two of the nine items: almost all students accurately associated weather with snowfall during winter storms (90%) and a majority associated climate with changing bird migrations (77%). Students scored less than 75% on the remainder of the items. Students had the most difficulty with two statements: "c. a summer heat wave with very hot temperatures" (38%) and "several decades with the most hurricanes ever recorded" (47%).

Following the unit of instruction (t_2), 75.2% students correctly identified weather or climate prompts. The significant increase in posttest scores indicated instruction was able to improve student distinction between weather and climate. At the end of instruction, while scores increased on seven of the nine items, students only made significant gains ($p< 0.05$) on three statements, "a summer heat wave with very hot temperatures" (which they struggled with on the pretest (t_1)), "a major outbreak of tornadoes with loss of life" and "a summer season with the most hurricanes ever recorded." Instruction was not able to significantly alter student perceptions on the "increase in hurricanes over several decades." Largest gains were made on the summer heat wave statement. Normalized gain and effect size calculations suggest modest or medium conceptual gains from pre to posttest (Figure 1).

After six months (t_3), student perceptions of weather and climate continued to increase significantly ($p<0.05$). Improvements were measured between post instruction (t_2) and six months after instruction (t_3) ($p < 0.05$) on two statements: drying up of a large lake*,* a ten-year period with the most hurricanes ever recorded. Normalized gains and effect size suggest medium conceptual gains in the time between instruction and the delayed posttest (t_3) (Figure 1).

Table 1

Percentage (%) and Frequency (n) of Correct Responses

Weather-climate prompt (correct response)	**Pretest % items correct (n)**	**Post test % items correct (n)[1]**	**Delayed post % items correct (n)[2]**	**Pretest to delayed post[3] % change**
a) drying up of a large lake *(climate)*	57% (27)	68% (32)	89% (42)*	32%*
b) a winter storm that dumps a large amount of snow *(weather)*	89% (42)	94% (44)	97% (46)	8%
c) a summer heat wave with very hot temperatures *(weather)*	38% (18)	64% (30)*	72% (34)	34%*
d) leaves budding out on trees earlier and earlier in the spring *(climate)*	66% (31)	74% (35)	74% (35)	8%
e) a warmer winter without any major snowstorms *(weather)*	64% (30)	62% (29)	74% (35)	10%
f) a major outbreak of tornadoes with loss of life *(weather)*	72% (34)	96% (45)*	91% (43)	19%*
g) birds migrating to warmer areas later and later in the fall *(climate)*	77% (36)	87% (41)	89% (42)	12%
h) a summer season with the most hurricanes ever recorded *(weather)*	62% (29)	85% (40)*	83% (39)	21%*
i) a ten-year period with the most hurricanes ever recorded *(climate)*	47% (22)	47% (22)	85% (40)*	38%*
Category Averages (st dev)	**63.6% (19.2)**	**75.2%* (17.5)**	**85.2%* (17.0)**	**18%**

Note. *significant changes $p<0.05$, [1]pretest to post test, [2]post test to delayed posttest, [3]pretest to delayed posttest

Figure 1

Pre-, Post-, and Delayed Post-Test Means, Standard Deviations, Normalized Gains and Effect Sizes

Prompt Category	Statistics	PreTest	Pre to Post Change	Post Test	Change Post to Delayed Post	Delayed Post test	Change Pre to Delayed Post		Change
Weather (9 prompts)	Mean	63.6%		75.2%*		84.2%			
	Std Dev	19.2		17.5		17.0			
			0.32		0.36		0.56		Norm. Gain
			0.60		0.51		1.1		Effect Size
Greenhouse Gases (7 prompts)	Mean	52.9%		62.6%		58.7%			
	Std Dev	15.5		15.3		18.4			
			0.21		-0.11		0.12		Norm. Gain
			0.63		-0.21		0.31		Effect Size
Human Actions (6 prompts)	Mean	39.7%		55.0%		51.4%			
	Std Dev	29.6		24.8		25.2			
			0.25		-0.08		0.19		Norm. Gain
			0.52		-0.14		0.44		Effect Size
Overall	Mean	48.8%		59.7%		59.7%			
	Std Dev	10.8		11.7		11.1			
			0.21		0.0		0.21		Norm. Gain
			1.0		0.01		1.0		Effect Size

Note. *significant change $p<0.05$.

Qualitative interviews offered a look into the benefit of students talking in the presence of one another. For example, one student when responding to the statement "the climate where I live changes from day to day" emphatically stated "I strongly disagree because climate doesn't change from day to day." Another student agreed and replied, "Because it takes like 30 years to change climate." Discussions such as this one offer students the opportunity to add more detail to an already correct response and simultaneously reveal student thinking.

Greenhouse Effect & GHG

Question 15 asks students about the benefits of the greenhouse effect. Only 13% of students recognized the importance of the greenhouse effect and its benefits to humans prior to instruction (t_1). This significantly increased ($p<0.001$) to 49% following instruction (t_2). Unfortunately, students reverted back to their prior understanding six months later (t_3) when only 19% answered this prompt correctly, a significant decrease ($p< 0.001$). Almost 50% of the students indicated on the delayed posttest (t_3) that the greenhouse effect was harmful to the Earth. This appears to contradict the student responses during the interviews when students were asked about living on a planet with greenhouse gases. The students overall indicated they would need the greenhouse gases to stay warm. "I'd like to live on a planet with greenhouse gases because they keep it heated. Without them it would just be cold, and we couldn't live on a planet that didn't have them?"

Item 16 asks students to identify specific greenhouse gases. Carbon dioxide was identified as the main greenhouse gas by only 51% of the students on the pretest (Question 16: from a choice of only oxygen, nitrogen and CO_2). Students were significantly ($p<0.05$) more likely to identify CO_2 as the correct GHG of the three provided following instruction (68%) with no significant change six months later (t_3).

Question 17 asked students to identify greenhouse gases from a list including oxygen, carbon dioxide, nitrogen, methane, water vapor, argon, and nitrous oxide. Prior to instruction (t_1) students were able to identify carbon dioxide as a greenhouse gas 81% of the time. They were also able to indicate correctly that argon is not a greenhouse gas 76% of the time. Following instruction (t_2) students were still correctly identifying carbon dioxide and argon. Students made significant gains in their ability to identify examples and nonexamples of GHG following instruction including water vapor ($p<0.001$), carbon dioxide ($p<0.01$), oxygen ($p<.01$), and methane ($p<0.05$). In support of this, students consistently identified carbon dioxide as a GHG during interviews (t_3). Students in each interview group also recognized that methane contributed as a GHG, but the explicit identification of water and nitrous oxide was missing. On the posttest (t_2), 49% of students correctly identified nitrous oxide as a GHG but no progress was made. There was no change in the perception of students who incorrectly identified nitrogen (23%) or argon (77%) as GHG.

Curiously, six months later (t_3) students were less likely to identify carbon dioxide ($p<0.01$) and water vapor ($p<0.01$) as GHG, reverting back to their pretest understanding, but students were more likely to correctly identify nitrous oxide ($p<0.05$) as a GHG. As indicated in the interviews some of the correct answers could be attributed to logical guessing. We use this term to identify an answer choice that is not completely understood by the student but does connect to something that they remember. Unit instruction was able to improve student understanding of GHG, specifically carbon dioxide, methane, water vapor, and nitrous oxide. Overall the lesson was not successful in changing preconceptions over the long term since the scores for carbon dioxide, nitrogen, and argon reverted back to pretest levels. During semi-structured interviews, students were asked how greenhouse gases impact our planet. While student groups all identified them as something that warms our planet, only rarely did they suggest that they were a benefit. Once prompted, students responded that they were beneficial to the planet, but the fact did not arise when asked how they impact the planet. Students may be transferring the negativity associated with increasing GHG, forgetting they are needed to support life on our planet. All student groups made a connection between the ozone layer and the greenhouse gases. Including statements like: "It makes holes in the ozone layer" and "It hurts the ozone layer."

Human Actions & Greenhouse Gases

Item 19 assessed students understanding the effects of six human actions on greenhouse gases. Before instruction (t_1), 68% of students attributed GHG to a combination of human and natural sources. Although not significant, instruction was able to increase that percentage to 79% (t_2) but six months later the number of correct responses dropped lower than the pretest to 60% (t_3). More students were likely to attribute GHG production to burning of fossil fuels on the delayed posttest (t_3) than on the pretest (t_1). This is not surprising given the discussions in the interviews. Many students were consistently reluctant to acknowledge the role humans play in production of GHG. For example, in one group, when students were asked "Are greenhouse gases more from natural sources or more from man-made sources? Every student responded, "Natural." In another group when asked: "…do you believe that humans are causing our climate to change, presently? One student replied, "I think we're having an impact on it but it's not completely on us." Again, all students agreed with this idea, "Yeah, like I think it is sort of us, like I think we do have an impact on it, but it's not all of us doing it. It's happening naturally too." Students consistently expressed this idea. For example, another student in another group said, "I just, I mean it's been going on like this for a long time. Stuff's been getting hotter and stuff's been getting colder. That's just how things roll" and most students in the group agreed.

Additionally, understanding climate change includes the knowledge of impacts of human actions on the amount of GHG in the atmosphere which was addressed in question 20. Students were required to identify how each of the six human action prompts would impact the amount of GHG. Prior to instruction (t_1), 50% or higher of the students could correctly identify impacts of burning fossil fuels, planting trees, and using alternative energy sources (Table 2). Students significantly improved on all of these ($p<0.05$) on the posttest (t_2). Almost 70% of students recognized the impact of driving automobiles prior to instruction with no significant change afterward. Students had the most difficulty determining relationships between GHG and a c) expanding the size of the ozone hole or e) insulating buildings both before and after instruction. Six months after instruction (t_3) significant gains occurred for both of these items ($p<0.05$). This initial struggle was not surprising given the confusion, discussed above, that students have understanding the relationship between the ozone layer and GHG.

Table 2

Percentage (%) and Frequencies (n) of Correct Responses

Human Actions prompts	**Pretest % correct (n)**	**Post test % correct[1] (n)**	**Delayed Post % correct[2] (n)**	**Pre to Delayed post % change[3]**
a) burning of oil or coal for fuel	55% (26)	83% (39)*	68% (32)	13%
b) planting trees and forests	47% (22)	66% (31)*	62% (29)	15%
c) expanding the size of the ozone hole	4% (2)	17% (8)	19% (9)	15%*
d) using alternative energy sources such as solar power and wind	53% (25)	68% (32)*	60% (28)	7%
e) insulating buildings to prevent heat loss/gain	11% (5)	17% (8)	32% (15)	21%*
f) driving automobiles	68% (32)	79% (37)	68% (32)	0%
Category averages (Std Dev)	**39.7% (29.6)**	**55.0%* (24.8)**	**51.4%* (25.2)**	**11.8%***

Note. *significant changes $p<0.05$, [1]pretest to post test, [2]post test to delayed posttest, [3]pretest to delayed posttest

Discussion

Our first research question focused on Appalachian middle school students' perceptions of weather, climate, and climate change. Results of this study indicated that the students acknowledged that the Earth's climate changes. Students exhibited increased understanding about basic concepts concerning weather and climate using the length of time as the distinguishing factor. Students were clear about how greenhouses work and the necessity of GHGs to moderate temperature on earth. Students attributed both natural and man-made factors to global warming. But digging deeper into all these general ideas, revealed many areas where students continue to wrestle with accurate scientific content.

Our second research question targets the persistence of these understandings and how they may or may not have changed after six months without further targeted instruction. Results for the greenhouse effect and GHG showed increased students' understanding but these increases are short-lived, essentially disappearing six months post instruction (t_3). Students particularly struggled with recognizing the benefits of the greenhouse effect. Instruction was able to alter this understanding, but this gain was lost over time. We choose to look at these results as evidence that students need increased time and consistent exposure to successfully develop the complex understandings needed to both understand these ideas and to mesh them with common worldview ideas. These common worldview ideas are often scientifically inaccurate, but more likely to be personally comfortable (Clifford & Travis, 2018). Some of the critical questions regarding the benefits of the greenhouse effect and the sources of the GHG were quite troubling because of the decay in understanding over time. In fact, more students were likely to attribute GHG production to only the burning of fossil fuels on the delayed posttest (t_3) than on the pretest (t_1). These results, while disappointing, are not surprising given the complicated nature of the greenhouse effect and the low percentage of adults that correctly understand the greenhouse effect.

During the semi-structured interviews, students consistently identified "time" as the critical factor in distinguishing between climate and weather effects. References to daily and weekly phenomena representing weather were consistent. Sample statements included "weather is like a weekly thing," "weather is more daily than weekly" and "it takes 20- and 30-year period for a climate to change." While students did not express the exact same understandings about weather and climate (i.e., weekly vs. daily) they were clear that weather was differentiated from climate by lengths of time. However, when asked to apply that criteria to weather- and climate-influenced events, pretest responses revealed many ideas that were scientifically inaccurate. These results are consistent with previous research findings (Spiropoulou et. al. 1999; Read et. al. 1994; Gowda, Fox, & Magelky, 1997; Papadimitriou, 2004; Lombardi & Sinatra, 2012) and offer us recommendations for teaching in the future which are summarized in the Conclusions.

A closer look at the students' responses during the interviews reveals important sense-making and indicates complexity behind students' choice of words (Prain, 2006). For example, during the interview Ashley responds to a question about the weather where she lives with the following, "...but here it's like we have mild winters and not very hot summers. It means that the *climate* changes a lot." Ashley's use of the word climate could be interpreted as simply "incorrect" or it can be viewed as a sensible interchange of the terms weather and climate. Students invoke terms that make sense to them, in the vernacular or everyday, without considering the scientific meaning. In this way we see Ashley's response as sensible but scientifically inaccurate. This subtle difference in interpretation matters because it offers teachers a different way to engage with students beyond simply identifying their inaccuracies and correcting them. Teachers can open a discussion around "everyday" use of terms and "scientific" use of these terms which can assist students in developing more sophisticated scientific understandings (Hammer & van Zee, 2006).

Additionally, we compared our students' delayed post scores with the previously published research with similarly aged students. Since the assessment measure was built upon previous published research, it is easy to compare these results with Boon's (2009) study which included students' understanding of these issues in two different decades and locations: UK students in 1991 and Australian students in 2001. Only 26% of Appalachian middle schoolers knew that the sea level would rise with a warmer climate, while 76% of Australian and 66% of UK students knew this. This trend was observed in the interviews where students were clearly still working through the outcomes of global warming on sea levels. The outcome that students, regardless of country or decade, knew most accurately was that polar ice caps would melt under the influence of a warmer climate (83% Appalachian, 85% UK, and 85% Aus). This was the only question in which there was no significant difference across all comparisons. Again, given the relatively straightforward connection between increased temperature and melting ice this makes sense.

Another outcome question that was asked of these students was the following: How might these actions impact the GHG in our atmosphere? This question requires students to evaluate certain actions that they could take and determine if they might increase or decrease the amount of GHG in our atmosphere. Interestingly, all 3 groups of students did similarly well on 1 action: using alternative energy. Approximately, 60% of all three groups of students knew that these actions would decrease the amount of GHG in our atmosphere. The groups answered differently to the following two actions: burning oil or coal (68% Appalachian, 80% UK, and 83% Aus) and driving automobiles (68% Appalachian, 76% UK, and 84% Aus). Interestingly, the greatest source of anthropogenic carbon dioxide (burning fossil fuels) was only accurately identified by just 68% of Appalachian students which live in an important coal source region of the country.

Although the studies were slightly different, comparisons can also be made with the students' assessed from an urban area in the NE (Bodzin et al., 2014) at roughly the same time period (2011-2012). Both groups of students held basic understandings of the differences between weather and climate but were unable to apply this rudimentary understanding when considering the complex interactions between weather and climate and the timescales associated with changes in climate. Bodzin's students struggled more with identifying the appropriate GHG and not recognizing the importance of water vapor as a GHG (only 23.3% of the Bodzin students correctly identified the 3 gases -- carbon dioxide, water vapor, and methane-- provided on their multiple choice question) yet the students in this study were more successful: 62% selected carbon dioxide, 65% water vapor and 55% methane as a GHG.

Bodzin et al. (2014) also asked students to provide types of human activities that are causing long-term increase in carbon dioxide levels whereas our study asked students to categorize certain actions and how they might impact the amount of GHG in the atmosphere. The Bodzin study reported that 61.2% of students provided adequate responses that were vague but accurate (including transportation use, using more heat in the winter and air conditioning in the summer, burning fossil fuels, etc). The current study found the following accurate associations between actions and impacts on GHG amounts in the atmosphere: 68% for burning oil or coal for fuel (*increase GHG)*; 62% for planting trees and forests (*decrease GHG*), 19% for increasing the ozone hole (*does not impact GHG)*, 60% for using alternative energy sources (*decrease GHG*), and 68% for driving automobiles (*increase GHG*). So, in general with the exclusion of the expected confusion of the ozone hole, approximately 60% of students could accurately categorize these actions which was quite comparable to the 61.2% of students with adequate responses in the Bodzin et al., (2014) study. This finding also opens up an opportunity to invoke student sense making in the research analysis. As previously mentioned, students often conflate the hole in the ozone layer with GHGs and climate change. On the surface this may just seem incorrect, but it actually makes sense if we look at the ideas behind the two phenomena. Both phenomena (the ozone hole and climate change) have to do with atmospheric processes, and both are considered environmental issues. It should not surprise teachers then that students conflate and confuse the two. We suggest that by helping students see that these two phenomena *are* related, students can develop deeper understandings of both content ideas.

The 5E model used as an instructional tool for this study, builds on the ideas of Driver et. al. (2014) and Hammer and van Zee (2006) in that the instructor was intentional about incorporating students' lived experiences and ideas into the explanation and evaluation phases of the unit. In this way students' experiences became a part of the curriculum. Even when students expressed "misconceptions" the teacher was open to seeing how these developed and how they could be extended toward more "correct" and "scientific" explanations. The persistent increase in correct responses for several questions and topics would then suggest that this method of instruction shows promise for these topics. Given that the increases in correct responses did decay over time for some areas, additional activities and/or extending the time for instruction are recommended.

Limitations and Future Work

Several limitations should be noted particularly the small sample size and the convenience sampling technique. Researchers were only able to provide the unit of instruction to a school within driving distance from the university. Other classrooms were contacted but this classroom and teacher were willing to work with the researcher who delivered the unit of instruction. The limited sampling size does call to question the generalizability of the results, but the students do represent a cross-section of the broader school community. Additionally, interviews with students were only conducted once at the delayed post instruction time frame (t_3). Conducting interviews with students at each time frame would allow a deeper discussion on student content retention, persistence of alternative conceptions, and how learners construct knowledge.

Since this study, the state has adopted an "adapted" version of the Next Generation Science Standards. In fact, this adaptation was a direct result of the political climate when a member of the state Board of Education changed the wording for several standards related to climate change. For the middle school standard, the state board of education changed the standard "MS-ESS3-5: Ask questions to clarify evidence of the factors that have caused the rise in global temperatures over the past century" to "S.6.ESS.6 ask questions to clarify evidence of the factors that have caused the change in global temperatures over the past century." This could be considered as one piece of evidence of the "cultural bias" that makes these topics challenging to instruct, particularly in this area which is so strongly supported by the coal industry. We plan to establish this initial research as a baseline of student understanding prior to NGSS implementation. Additional studies will be conducted to determine whether simply adding the required content to the curriculum significantly improves student understanding. We plan to assess a new group of middle-level students who will be receiving instruction after the adoption of the NGSS standards for this state and compare these two groups of students 1) at the middle-level and again 2) at the secondary-level when the standards will more fully align with the assessment measure.

Conclusions

We know from research that students struggle to understand the long-term idea of climate vs. weather. While they do consistently include time as a factor that differentiates the two, answering questions that are more outcome- and application-oriented remains challenging. Students seem to develop an understanding of weather as evidenced by the increase in score immediately after instruction for three weather items on the assessment measure. This makes sense if we consider that weather, on a daily basis, is what students experience first-hand themselves. Climate changes over decades of time are less likely to be understood by students as they are less acutely felt by them. The largest measured improvements in understanding were made in the statement regarding a heat wave in the summer. However, as evidenced by the delayed post (t_3), two climate related items, substantially improved over time. Recognizing this as an expected result for adolescents has implications for curriculum design in the future. A solid understanding of weather can potentially support a better understanding of climate. Recommending that teachers include specifics in instruction is a key take-away from this research. For example, while we know that students make the connection between melting polar ice caps and global warming, students did not make that connection to rising sea levels. Drawing more causal connections and specifically discussing how the changing temperature does, in fact contribute to rising sea levels would help students make this multi-step connection.

Instruction resulted in short-term improvements but only a few changes were still evident six months after GHG and greenhouse effect instruction. Not surprisingly, well documented areas of confusion remained unaltered. Students had the most difficulty determining relationships between GHG and expanding the size of the ozone hole or insulating buildings although this did significantly improve from pre (t_1) to 6 months post instruction (t_3). However still only 15% and 21% (respectively) answered these questions correctly. This is not surprising at all given the confusion, discussed above, that students have about the relationship between the ozone layer and GHG which has been well documented in prior research. Our work supports that which has been previously supported by research like Hammer and van Zee (2006), we should approach students' ideas as "common sense" because many times "incorrect" student ideas (commonly called misconceptions) are actually quite sensible given students' experiences in the world, particularly as set within the cultural context of the local community. Unless as teachers, we understand where students are coming from in their thinking, and this includes conceptual and cultural sense-making, it is highly unlikely that the foundations of students' conceptual ideas will be open to long-term change (Von Glasersfeld, 2013).

Using a constructivist theoretical framework and expecting that students will try to make sense of the questions they are asked, can support teachers to reframe instructional approaches. If teachers assume that students' responses are not just "wrong," but rather conceptually incomplete and often, sensible, teachers can view students' learning in a more productive way. Given this theoretical lens the results presented here and elsewhere are not surprising and offer science educators a way to see logical sense making in many of the ideas that students revealed. We share here a quote by Driver at al. that captures the nature of the way that we see the student responses.

Pupils come to science lessons with ideas about the natural world. Effective science teaching takes account of these ideas and provides activities which enable pupils to make the journey from their current understandings to a more scientific view. (Driver, Squires, Rushworth, & Wood-Robinson, 2014, p. xiv).

In their book on secondary students' understanding of science concepts (2014), Driver and her colleagues stress how understanding where students are in their scientific thinking and recognizing where that thinking comes from, are critical to helping students on their journey to a more scientific view of the world. Despite being grounded in student sense-making, and helping teachers think about how to teach science, constructivism was, for many years, used as a way to dismiss student ideas as incorrect and a justification for those ideas to be completely changed as a part of instruction (Osborne, 1996). In contrast to this, Driver and colleagues suggest that we see student understandings as a starting point and a place to work from. It is critical that we recognize that experiences that students have in their real world lives inform the scientific understandings that they eventually develop. Instead of fighting with students' emergent ideas, educators can embrace them and use them as a launching pad for more sophisticated learning. We operate from previously mentioned Hammer and Van Zee (2006) frame of mind as we look at the challenges and opportunities the students in this study afford us as educators looking to understand students' ideas on climate, weather and the greenhouse effect.

The publication of the NGSS document has helped address the importance of climate change inclusion throughout K12 education. Prior to the adaptation and adoption of the NGSS in this Appalachian region, climate change was only addressed in high school environmental science and advanced earth science electives. Weather was taught in elementary and middle school with little mention of climate. Now with NGSS performance expectations driving the curriculum, students are introduced to weather as early as kindergarten and exposed to climate in third grade. Additionally, the weather and climate science topics are reinforced in fifth grade, middle school, and high school. This would suggest that states that adopt NGSS should have more structured reinforcement of concepts throughout their education perhaps avoiding students' reverting to previously held misconceptions about climate change. Those states who do not adopt NGSS should scaffold weather and climate standards throughout elementary, middle, and high school to ensure gains in climate science understanding is not lost for lack of engagement.

This material is based upon work supported by the National Science Foundation under Grant OIA-1458952.

Tina J. Cartwright (Tina.cartwright@marshall.edu) is a professor of science education at Marshall University. Her research focuses on student conceptions of weather vs. climate and how that may impact their perception of climate change. She also teaches elementary science and evaluates the impact of utilizing the after-school setting to improve students' self-efficacy in teaching science.

Deb Hemler (dhemler@fairmontstate.edu) is the Chair of the Natural Sciences Department at Fairmont State University. She is the Director of the Education Resource Center at the Katherine Johnson NASA IV & V Facility in Fairmont, West Virginia. Her current research involves evaluating the efficacy of geoscience professional development for K12 educators.

Paula A. Magee (pamagee@iupui.edu) is a Clinical Professor of Science Education in the School of Education at Indiana University—Purdue University Indianapolis. Her research focuses on equitable and anti-racist science teaching in K12 schools and in teacher preparation programs.

References

Bofferding, L. & Kloser, M. (2015). Middle and high school students' conceptions of climate change mitigation and adaptation strategies. *Environmental Education Research*, 21(2), 275-294. http://doi.org/10.1080/13504622.2014.888401

Bodzin, A. M., & Fu, Q. (2013). The Effectiveness of the Geospatial Curriculum Approach on Urban Middle-Level Students' Climate Change Understandings. *Journal of Science Education and Technology*, 1–16. http://doi.org/10.1007/s10956-013-9478-0

Bodzin, A. M., Anastasio, D., Sahagian, D., Peffer, T., Dempsey, C., & Steelman, R. (2014). Investigating Climate Change Understandings of Urban Middle-Level Students. *Journal of Geoscience Education*, 62(3), 417-430.

Boon, Helen (2009) *Climate change? When? Where? Australian Educational Researcher*, 36 (3). pp. 43-65.

Bostrom, A., M. G. Morgan, B. Fischhoff, & D. Read. (1994). What do people know about global climate change? Mental models. *Risk Analysis,* 14 (6), 959–970. http://doi.org/10.1111/j.1539-6924.1994.tb00065.x

Boyes, E., Chuckran, D., & Stanisstreet, M. (1993). How do high school students perceive global climate change: What are its manifestations? What are its origins? What corrective action can be taken? *Journal of Science Education and Technology*, 2, 541–557.

Boyes, E., & Stanisstreet, M. (1993). The "greenhouse effect"; children's perceptions of causes, consequences and cures. *International Journal of Science Education*, 15, 531-552.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins and effectiveness. *Colorado Springs, Co: BSCS*, *5*, 88-98.

Carr, M., Barker, M., Bell, B., Biddulph, F., Jones, A., Kirkwood, V., Pearson, J. and Symington, D. (2013). The constructivist paradigm and some implications for science content and pedagogy. In The content of science: A constructivist approach to its teaching and learning (pp. 159-172). Routledge.

Clifford, K., & Travis, W. (2018). Knowing climate as a social-ecological-atmospheric construct. *Global Environmental Change*, 49, 1-9.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. London: Routledge

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.) Upper Saddle River, NJ: Pearson Education.

Dawson, V. (2015). Western Australian High School Students' Understandings about the Socioscientific Issue of Climate Change. *International Journal of Science Education*, *37*(7), 1024–1043. http://doi.org/10.1080/09500693.2015.1015181

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (2014). *Making sense of secondary science: Research into children's ideas*. Routledge, London.

Fisher, B. (1998). Australian student's appreciation of the greenhouse effect and the ozone hole. Australian Science Teachers Journal, 44(3), 46-55.

Gowda, M. V. R., Fox, J. C., & Magelky, R. D. (1997). Students' understanding of climate change: insights for scientists and educators. *Bulletin of the American Meteorological Society,* 78, 2232–2240.

Hammer, D., & van Zee, E. (2006). *Seeing the science in children's thinking: Case studies of student inquiry in physical science*. Portsmouth, NH: Heinemann Educational Books.

Harker-Schuch, I., & Bugge-Hendrikson, C. (2013). Opinions and knowledge about climate change science in high school students. AMBIO. 42:755-766.

Jarrett, L., Ferry, B., & Takacs, G. (2012). Development and validation of a concept inventory for introductory-level climate change science. *International Journal of Innovation in Science and Mathematics Education*, 20(2), 25–41.

Kılınc¸ A, Boyes E, Stanisstreet E (2011) Turkish school students and global warming: beliefs and willingness to act. *Eurasian Journal of Mathematics & Science Technology Educatio*n, 7(2), 121–134.

Klosterman, M., & Sadler, T. (2010). Multi-level assessment of scientific content knowledge gains associated with socioscientific issues-based instruction. *International Journal of Science Education,* 32(8), 1017–1043.

Libarkin, J. C., Anderson, S. W., Science, J. D., Beilfuss, M., & Boone, W. (2005). Qualitative analysis of college students' ideas interviews and open-ended questionnaires. *Journal of Geoscience Education*, 53(1), 17–26.

Lombardi, D. & Sinatra G.M. (2012). College students' perceptions about the plausibility of human-induced climate change. *Research in Science Education*, 42(2), 201–217.

Lombardi, D., Brandt,C., Bickel, E. & Burg, C. (2016). Students' evaluations about climate change. International Journal of Science Education, 38(8), 1392-1414.

McNeill, K. L., & Vaughn, M. H. (2012). Urban High School Students' Critical Science Agency: Conceptual Understandings and Environmental Actions Around Climate Change. Research in Science Education, 42(2), 373-399. doi:10.1007/s11165-010-9202-5.

National Climatic Data Center. (2008). *Weather/climate events*. Retrieved April 21, 2008, from http://www.ncdc.noaa.gov/oa/climateresearch.html

Osborne, J. F. (1996). Beyond constructivism. *Science education*, *80*(1), 53-82.

Papadimitriou, V. (2004). Prospective primary teachers' understanding of climate change, greenhouse effect, and ozone layer depletion. *Journal of Science Education and Technology*, 13, 299–307.

Prain, V. (2006). Learning from writing in secondary science: Some theoretical and practical implications. *International Journal of Science Education*, *28*(2-3), 179-201.

Prather, E. (2005). Students' beliefs about the role of atoms in radioactive decay and half-life. *Journal of Geoscience Education,* 53, 345–354.

Read, D., Bostrom, A., Morgan, M. G., Fischhoff, B., & Smuts, T. (1994). What do people know about global climate change? 2 Survey studies of educated laypeople. *Risk Analysis*, 14, 971–982.

Rye, J.A., Rubba, P.A., Wiesenmayer, R.L. (1997) An investigation of middle school students' alternative conceptions of global warming. *International Journal of Science Education,* 19(5), 527–551.

Shepardson, D.P., Roychoudhury, A., Hirsch, A., Niyogi, D., Top, S.M. (2014) When the atmosphere warms it rains and ice melts: seventh grade students' conceptions of a climate system, *Environmental Education Research*, 20(3), 333-353.

Shepardson D, Niyogi D, Choi S, Charusombat U. (2009). Seventh grade students' conceptions of global warming and climate change. *Environmental Education Research*. 15(5):549–570.

Spiropoulou, D., Kostopoulos, D., & Jacovides, C.P. (1999). Greek children's alternative conceptions on weather and climate. *School Science Review*, 18(294), 55-59.

Varma K. & Linn M.C. (2012). Using interactive technology to support students' understanding of the greenhouse effect and global warming. *Journal of Science Education and Technology*, 21(4), 453–464.

Von Glasersfeld, E. (2013). Radical constructivism (Vol. 6). London: Routledge.

Appendix A: Assessment Measure

1. Do you agree to participate in this research study? By selecting Yes, you agree to participate in this research study, which is collecting information from students around the world on their understanding and feelings about the greenhouse effect and climate change. You will not be asked to provide any personally identifiable information, and your participation is completely voluntary. You do not have to answer any questions that may make you feel uncomfortable. At any point in the questionnaire, you may choose to not participate and not submit your questionnaire.

- Yes
- No

2. What is the name of the city where your school is located?

3. What is the name of the state where your school is located? If you live outside the United States, please include your country as well.

4. If you know it or your teacher provides it to you, what is the latitude of your school?

5. If you know it or your teacher provides it to you, what is the best description of your climate zone (according to the Koppen-Geiger climate classification)?

- A – Equatorial
- B – Arid
- C - Warm Temperate
- D – Snow
- E - Polar

6. Think of the weather and climate you have experienced in the last few weeks. Has it been warmer, colder, or the same as a typical season?

- Warmer
- Colder
- the same

7. Which of the following is best explained by a change in the weather or the climate? (Answer choices are "weather" or "climate")

7a. drying up of large lake over man years
7b. a winter storm that dumps a large amount of snow
7c. a summer heat wave with extremely hot temperatures
7d. leaves budding out on trees earlier and earlier in the spring over many years
7e. a warmer winter without any major snow storms
7f. a major outbreak of tornadoes with loss of life
7g. birds migrating to warmer areas later and later in the fall over many years
7h. a summer season with the most hurricanes ever recorded
7i. a ten year period with the most hurricanes ever recorded

8. Does Earth's climate change?

- Yes
- No

9. What are possible causes or factors that might contribute to the change (or stability--lack of change) in the earth's climate?
10. Have you ever been in a greenhouse on a warm summer's day?

- Yes
- No

11. Do you think it is warmer or cooler inside a greenhouse than outside?

- ***warmer
- cooler
- the same
- I don't know

12. Explain why this might be so

13. What do you think the "greenhouse effect" is?

14. How might the "greenhouse effect" impact the earth's climate?

15. The "greenhouse effect"...

- ***benefits humans
- harms our earth
- does nothing to humans or earth
- I do not know

16. Which of the following do you think is the main "GHG"?

- Oxygen
- ***Carbon Dioxide
- Nitrogen

17. Which of the following are "GHG," if any? (You can choose more than one response.)

- Oxygen
- ***Carbon Dioxide
- Nitrogen
- ***Methane
- ***Water vapor
- Argon
- ***Nitrous Oxide

18. What do you think the outcomes of a warmer climate will be?
(answer choices include “different locations will change differently”, “significantly decrease”, “decrease”, “will not change”, “increase”, “significantly increase”)
18a. sea levels will…
18b. rainfall will…
18c. sunshine will..
18d. farmers crops will be….
18e. the ice caps in the North and South Poles will…

19. GHG originate...

- entirely from human activity.
- entirely from natural sources.
- entirely from fossil fuels.
- ***from a combination of human and natural sources.

20. How might these actions below impact the amount of GHG in our atmosphere?
(answer choices are traditional 5 point likert -- significantly increase to significantly decrease)
20a. burning oil or coal for fuel
20b. planting trees and forests
20c. expanding the size of the ozone hole
20d. using alternative energy sources such as solar power and wind
20e. insulating buildings to prevent heat loss/gain
20f. driving automobiles

Appendix B: Semi-structured Group Interview Questions

1. What's the difference between weather & climate?

2. Weather vs. climate – Do you agree/disagree with the following statements:

    a. The clothes that people wear are influenced by the weather.

    b. The weather where I live changes dramatically.

    c. The climate where I live changes from day to day.

    d. Changes in the weather means that the climate will change.

    e. The clothes available to buy in my local stores are influenced by the local climate.

3. What are the greenhouse gases? How do the greenhouse gases impact our planet? Where do greenhouse gases come from? Would you like to live on a planet with greenhouse gases in the atmosphere?

4. What was the weather like this past winter? How did it compare to "normal"?

5. What has the weather been like this spring?

6. How do you think our weather is related to what people call "global warming"?

7. How do you think "global warming" is related to climate change?

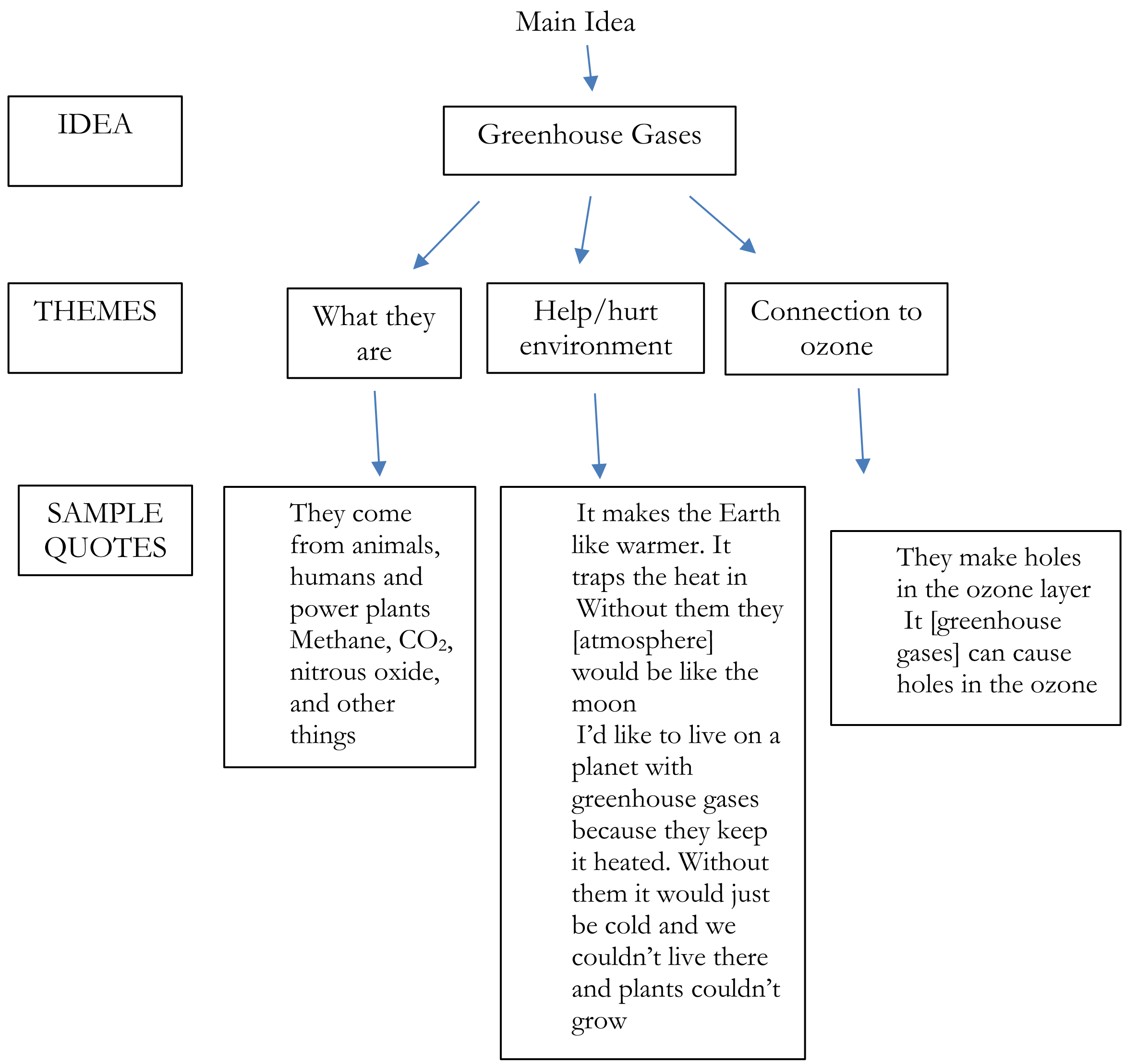
8. What do you think the outcomes of a warmer climate will be? How would our planet be changed if it became warmer?

    a. Relate to changing sea levels, farm crops, rainfall, sunshine, ice caps

9. Do you think the earth's climate can change? If so, what are possible causes that may lead to that change?

10. Do you believe that humans are causing the climate to change?

Appendix C: Coding Scheme Example



Appendix D: Unit Summary

Summary of the unit of instruction with the State Standard (valid in 2011), corresponding Essential Question, student learning objective, and 5E lesson cycle summary and question numbers that corresponded to this topic in the student questionnaire.

Day	Standard	Essential Question	Learning Objective	Summary	Q
1-2	SC.O.8.2.28 determine the impact of oceans on weather and climate; relate global patterns of atmospheric movement on local weather.	What is the relationship between weather & climate?	Students will differentiate between weather & climate.	The lesson introduces students to the ideas of weather and climate and how these two ideas are similar but yet different. During a PowerPoint discussion, students will utilize student response clickers to evaluate their understanding. In Elaboration, student examine the GLOBE Program global maps which demonstrate the factors which control Earth's climate.	7
3-4	SC.O.7.2.32 explain how changing latitude affects climate.	What causes the seasons?	Students will evaluate the influence of latitude on climate.	Students conduct a lab experiment using infrared thermometers and a table top globe that is situated 20 cm from a 100 watt bulb light source to determine the impact of the tilt of the axis of rotation on the surface temperature. In Elaboration, students reexamine the GLOBE Program maps of insolation and surface temperature across different months to find a pattern.	NA
5-6	SC.O.8.2.32 explain phenomena associated with motions in sun-earth-moon system (e.g., eclipses, tides, or seasons).	How's does the Earth's rotation & revolution impact the seasons?	Students will examine evidence of the Earth's rotation and revolution on the seasons.	Students consider evidence that they have collected (in their daily lives and also in the previous lesson) which can be used to support ideas related to Earth's rotation and Earth's revolution around the sun. Several movies are shown to provide further evidence of this large scale interaction between the Earth and the Sun.	NA
7-8	SC.O.7.2.27 examine the effects of the sun's energy on oceans and weather (e.g., air masses, or convection currents).	How does specific heat impact the Earth's climate?	Students will describe the influence of Earth materials on local climate.	Students collect data using heating lamps, thermometers and different earth materials to see how they respond to being heated and cooled. In Elaboration, students were given monthly average temperatures and annual rainfall amounts to graph of two cities in North America which have significantly different weather patterns because of their location and proximity to the ocean.	9
9-10	SC.O.7.2.27 examine the effects of the sun's energy on oceans and weather (e.g., air masses, or convection currents).	How can GHG be considered both a friend and an enemy?	Students will examine evidence of the effects of greenhouse gases on the atmosphere.	Students measure changing temperatures inside two tennis ball cans with 5 cm of water and 1 can with Alka-Seltzer tablets to release Carbon Dioxide. Student discuss/examine the role of GHG in the heat balance of the atmosphere and the evidence of an enhanced greenhouse effect with increasing amounts of CO2. In Elaboration, students consider the gases in the atmosphere and the daytime/nighttime temperature extremes of Earth, the Moon, Mercury, Venus and Mars.	13, 14, 15, 19, 20

ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 25, NO. 2, 30-46

A Case Study of a Researcher-Practitioner Partnership in Teaching STEM+C to Rural Elementary Students

Stephanie D. C. Playton
Longwood University
Texas Tech University

Rebecca Hite
Texas Tech University

Paula K. Leach
Longwood University

ABSTRACT

Both computer science (CS) knowledge and workforce readiness skills (e.g., creativity, communication, collaboration, and critical thinking) have equally grown in national importance to fill the growing pipeline of CS careers. Various factors have contributed to CS job shortages, which include a lack of student instruction on, engagement with, interest in and awareness of CS and careers. Rural students, an underrepresented group, lack access to CS content and pedagogies (inquiry-based instruction) that facilitate knowledge, skills, and affect towards CS. Some states are addressing the lack of CS and workforce readiness skills through new policies integrating workforce readiness skills and CS standards into formal education, starting in elementary school. The change in policy to integrate CS into elementary education fostered a researcher-practitioner partnership between researchers and three teachers. A single illustrative case study investigated how 18 contact hours of a three-unit inquiry-based integrated Science, Technology, Engineering, and Mathematics with computer science (STEM+C) curriculum augmented 34 rural fourth-grade (10 year old) students' engagement with, interests and attitudes in STEM+C and increased their knowledge of CS careers and use of workforce readiness skills. Analyses indicated significantly positive gains in interests and attitudes in science for all students, with the greatest improvement for girls. High levels of engagement were observed and self-reported for all students, but workforce readiness skills varied across the learning units. Results suggest that inquiry-based learning opportunities that integrate STEM with CS can support primary level students' interests and attitudes in STEM and foster workforce ready skills among geographically underrepresented students.

Keywords: case study, computer science, engagement, inquiry-learning, researcher-practitioner partnership, rural, STEM, STEM+C, workforce readiness skills

Introduction

Workforce statistics continue to reflect a high need in science, technology, engineering, and mathematics (STEM) jobs (US Bureau of Labor Statistics [BLS], n.d.) as they are critical to the American economy and development of innovations (National Science Board, 2015). STEM occupations are growing much faster than other occupations (Noonan, 2017); computer science (CS) jobs comprising over half of these projected jobs (Code Advocacy Coalition, n.d.). In recognition of the growing need for CS workers, states are passing policies to implement CS standards into the K-12 curriculum (Code Advocacy Coalition, n.d.; Sawchuk, 2017). In addition, states like Virginia (where the study took place) have coupled these policies with parameters to ensure students acquire *workforce readiness skills*, meaning their K-12 educational experiences foster communication, collaboration, critical thinking, and creative thinking. These attributes are also known as the 4Cs, which represent learning and innovation skills necessary for success in work and life in the 21st century (Partnership for 21st Century Learning, 2015). The first two skills of communication and collaboration relate to proficiencies in relating information and working with peers, whereas the latter skills of critical thinking and creativity describe the sophisticated thinking needed to address emergent issues and global problems (National Education Association [NEA], 2012). Other research has used the 4C paradigm in relationship to workforce readiness in adolescent literacy (Ehren & Murza, 2010), therefore, STEM subjects are a logical place for CS integration and to engage students in utilizing and practicing workforce ready skills (DeJarnette, 2012).

A 2018 report on the *State of Computer Science Education* by the Computer Science Teacher Association (CSTA) and Code.org Advocacy Coalition, stated that adoption of K-12 CS standards requires all schools to offer CS guided by policies to increase access to CS (e.g. by allowing CS to count towards core graduation requirements). In order for schools to have CS in secondary spaces, primary schools must begin integrating CS into their curricula. Schools and educators are preparing for CS policy implementations, especially in the elementary schools, where it is expected to be integrated into the current STEM curriculum, despite STEM having the least amount of instructional time (DeJarnette, 2012). Even though these policies are in place, it does not mean they will be fully supported, especially in rural areas that have less access to resources needed to teach the current curriculum (Johnson & Strange, 2007; Monk, 2007).

Since STEM learning is a broad area that embodies many subject areas and teaching and learning strategies vary (Lamb et al., 2015), it is important for the authors of the study to explicitly state the definition of STEM education in Virginia which entails 'authentic learning experiences for all students with an interdisciplinary and applied approach where all fields connect in complex relationships' (VDOE, 2017, para. 1). Lamb et al. (2015, p. 411) has suggested that this view of STEM can provide elementary teachers an opportunity to integrate more cross-curriculum learning approaches to the subject areas they are already responsible for teaching that “requires less specialization and more ability to see across areas of interaction and the resultant complexity within the STEM disciplines.” However, for STEM and CS education (referred herein as *STEM+C*) to become an ordinary part of elementary instruction, teachers need support to implement curriculum that is engaging, inquiry-based, and STEM integrated within classrooms (DeJarnette, 2012). Therefore, the purpose of this research was to explore how an inquiry-based integrated STEM+C curriculum, developed within a researcher-practitioner partnership, augmented rural fourth-grade (10 years old) students’ engagement with, interests and attitudes in STEM+C, as well as increased their knowledge of CS careers and use of workforce readiness skills. To establish the needs and gaps that this study addresses, we review literature on researcher-practitioner partnerships, elements of enhancing elementary students’ engagement, interests, and attitudes towards STEM+C, and the improvements to workforce readiness (skills) that inquiry-based STEM+C experiences provide to primary-level learners. This review of the literature provides an understanding as to how collaborative partnerships, and the interventions they design, are actively improving American STEM education.

Researcher-Practitioner Partnerships

An approach to connect education theory to classroom practice is through higher education institution partnerships between researchers and teachers, co-designing STEM lessons (DeJarnette, 2012). Researcher-practitioner partnerships (RPPs) commit to solve practical problems, such as new instructional content, through collaboration (LeMahieu et al., 2017). Involving and supporting classroom teachers in developing new curriculum has been incredibly successful (Webb et al., 2017); and allows for avenues for research that is more useful to practitioners (Baker, 2003). Hence, the RPP was a useful avenue for conducting research on co-created classroom intervention, like integrated STEM+C learning through an elementary curriculum.

Student Engagement, Interest, and Attitudes

In tandem with employing best practices, the curriculum should be engaging for students (Ainley, 2012; The New Teacher Project [TNTP], 2018). Engagement is acknowledged as a construct difficult to both define and measure (Christenson et al., 2012), therefore, in the context of this study, engagement is defined as the degree at which a student positively or negatively attends to, or shows an interest in the completion and involvement in a specific classroom activity through constructs of behaviour, affects/emotion, and cognition. Hence, engagement is an important attribute for improving student learning (Trowler, 2010), especially when related to academic achievement from elementary school learning experiences (Ainley, 2012; TNTP, 2018). Therefore, content such as STEM+C activities, must contain specific strategies to foster engagement for elementary students such to kindle their interests in CS and STEM.

The nature of the empirical relationship between engagement and interest (Lam et al., 2012) is considered as interrelated, given that interest is often a trigger to engagement (Ainley, 2012). Interest, in this study, is defined as a motivational variable to foster desire for learning (Frenzel et al., 2010). The connection between interest and motivation is significant as research from Osborne et al. stated that "motivation offers important pointers to the kind of classroom environment and activities that might raise pupils' interest in studying school science" (2003, p. 1049). Further, these authors relate the importance of interest to fostering positive *attitudes* for science, a vital component of science education. Attitudes are an overall evaluation of stimulus objects that are influenced by affective, cognitive and behavioural information (Haddock & Maio, 2004). Ensuring that STEM+C learning experiences are interesting and engaging to students can lead to positive attitudes towards STEM (Christensen et al., 2015), even for the youngest of learners (like pre-school, see Leibham et al., 2013). In turn, early and rich STEM+C experiences may help mitigate known declines in STEM interest in middle and high school (George, 2000; Sadler et al., 2012) when students begin establishing their career beliefs in middle school (Kier et al., 2014; Skamp, 2007). Hence, developing even a nascent awareness of STEM careers is vital for student in the primary grades (Dorph et al., 2017).

Current studies suggest that if we provide students with access to STEM as early as elementary school, it not only increases their interest in pursuing STEM careers (Ball et al., 2017; littleBits, 2018; Tran, 2018), but also reduces inequalities in access and opportunities to learn STEM and develop workforce readiness skills, like problem-solving and communication (Sarama et al., 2018; Tran, 2018). As CS has an inherent technological component, it may play a unique role in engaging elementary students in STEM learning (Kurz et al., 2015) by increasing students' interests in STEM by connecting it to a curricular context (Lam et al., 2012; Li et al., 2010; littleBits, 2018). Further, STEM-based elementary interventions that employ technology can increase interests and attitudes in STEM while supporting workforce readiness skills if coupled with strong pedagogies like inquiry-based learning (Eccles & Wang, 2012; Lam et al., 2012; Li et al., 2010).

Inquiry-based Learning for Workforce Readiness Skills and Student Engagement

Inquiry-based pedagogies can also support student engagement (DeJarnette, 2012; Finn & Zimmer, 2012; Trowler, 2010) and is often implemented through group work where workforce readiness skills like collaboration and communication skills are emphasized (Tran, 2018); however, there is still a lack of frameworks that guide best practices for implementing STEM programs, even those implemented at the secondary level (Heil et al., 2013). Simply exposing students to one-day events in STEM has not provided positive results in effectively increasing positive perceptions and interest in STEM careers in the primary years (Kurz et al., 2015). However, providing integrative approaches for classroom-based STEM activities has shown significant increases in attitudes (Toma & Greca, 2017; Tran, 2018) and interest (Lamb et al., 2015), suggesting that a curriculum developed through an RPP that focuses on engaging, inquiry-based integrated STEM+C curriculum would be beneficial to primary level learners.

Importance of RPP in Enhancing American STEM Education

The National Science Foundation (2017) has recently shifted more funding to elementary STEM learning, providing greater opportunities for researching elementary STEM+C learning experiences. As a result, the growing literature on STEM+C began with research on the efficacy of short-term interventions like one-day events (Kurz et al., 2015) and classroom lessons (Ball et al., 2017), to longer-term interventions including entire learning units (Lamb et al., 2015; Li et al., 2010; Toma & Greca, 2017; Tran, 2018). Kurz et al. (2015) has suggested that STEM+C interventions have fallen short of connecting content relevancy to students and lack in use of inquiry-based learning methods. These criticisms are well taken as connecting content learning to STEM careers can provide a valuable context to learning experiences (Li et al., 2010) and technology can increase interest (Kurz et al., 2015) and mitigate geographic disparities.

While policy demands make their way into classrooms to drive instructional changes, many educational institutes will struggle especially in those that lack resources (i.e., rural schools). Rural schools often lack the expertise of content area specialists dedicated to integrate concepts effectively, as well as the expertise of someone with integrating CS and general resources (Johnson & Strange, 2007; Monk, 2007). Interventions often model those that would be difficult for rural schools to replicate, as they are often hours away from experts that are generally located in the urban and suburban areas (Kurz et al., 2015), but connecting students with professionals in the STEM and specifically CS fields is important (Li et al., 2010). Rural areas need to capitalize on technologies, such as video conferencing, to provide diverse exposures to STEM+C careers.

Models of research describe interventions that include more long-term exposures to STEM+C, but still show the need for teacher support (Guzey et al., 2016; Toma & Greca, 2017; Tran, 2018). Other models have shown that learning experiences in STEM can gain momentum in changing students' interests (Lamb et al., 2015); however, teachers need support in providing more hands-on, inquiry-based activities in mathematics and science (DeJarnette, 2012). Educators are also challenged with a lack of resources in general (Guzey et al., 2016; Kotok & Kryst, 2017; Li et al., 2010), suggesting rural, underrepresented students will fall further behind in educational experiences (Biriescu & Babaita, 2014) as STEM+C initiatives challenge educators responsible for these large populations of students (Sawchuk, 2017).

Purpose of Study

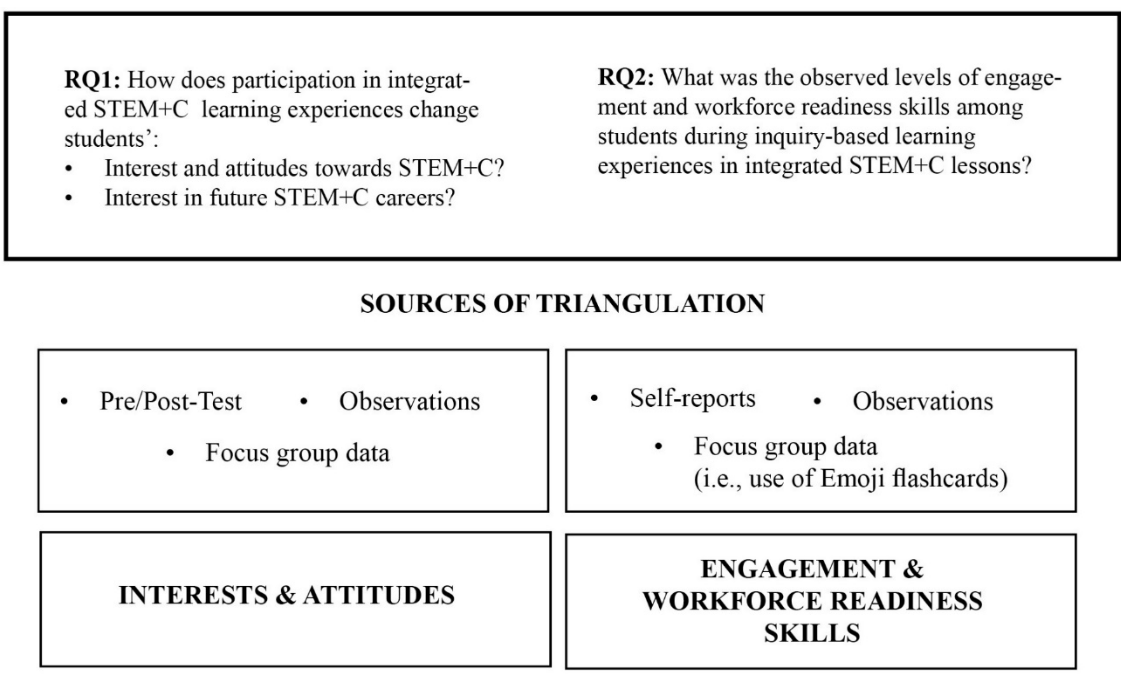
The purpose of this study was to examine how an integrated STEM+C curriculum for a primary-level audience influenced students' interests and attitudes in STEM, enhanced their knowledge of STEM careers, engagement and use of workforce readiness skills. A year-long (i.e., 18 contact hours) series of classroom-based interventions employed three central activities that emphasized use of workforce readiness skills to accomplish CS-related tasks (i.e., design/test a moving object, create sculptures with circuitry, and develop an ecosystem video game) with 34 rural fourth-grade students that were approximately 10 years of age. The questions and sub-questions that guided the research were:

1) How does participation in integrated STEM+C learning experiences change students':
   a. Interest and attitudes towards STEM+C?
   b. Interest in future STEM+C careers?
2) What was the observed levels of engagement and workforce readiness skills among students during inquiry-based learning experiences in integrated STEM+C lessons?

Methodology

The RPP was initially precipitated by requests from two teachers from a rural school who expressed concerns in new policies mandating the teaching of STEM+C. Researchers met with teachers, observed their classrooms, and provided feedback such to develop an 18-contact hour, 3-unit intervention for integrated learning experiences through inquiry-based pedagogies with a focus on STEM+C objectives. Therefore, a single illustrative case study design was selected since the unit of analysis (students) comprised of aggregate classrooms, who engaged with 3 separate units (activities) related to both the intervention (STEM+C) and constructs of interest (interest, attitudes towards STEM+C and related careers as well as engagement and use of workforce readiness skills). Multiple sources of evidence (surveys, self-reports, focus groups, observations) over a prolonged duration of time were collected for robustness (Yin, 2018). Figure 1 describes the triangulation across data sources, both qualitative and quantitative for case analysis.

Figure 1
Research Questions Aligned with Triangulated Data Sources



Thirty-four fourth graders from one rural elementary school in Central Virginia participated in the case study. The school reported that 54% of their student population for the 2017-2018 school year qualified for free and reduced lunch, indicating a majority of the school's population to be that of low socioeconomic status (SES). Students of low-SES tend to be represented less in STEM pathways (Barzanji, 2013; Niu, 2017; Xie et al., 2016), suggesting a viable sample. Therefore, the classroom intervention created and implemented within the RPP sought to facilitate engaged learning and cultivate positive interests and attitudes towards STEM+C by having under-represented fourth graders utilize workforce readiness (4Cs) skills. Data collection occurred over the course of four, small-group lessons. The lessons were part of larger learning units that included objectives for moving objects, electricity, and animal ecosystems (i.e., Data collection I, II, IIIa, and IIIb), respectively.

The first data collection occurred when students used probeware to measure friction while pushing or pulling an object designed to assist an animal to cross a busy road. As part of the process, students had to program a small robot to autonomously enter and exit the device they created. The second data collection took place during an electricity unit where students were challenged in pairs to use conductive and non-conductive playdough to create a sculpture that powered lights. Next, students had to use science terminology to verbally explain how they created their sculptures, then sequence the process of recreating their sculpture through a 'how to guide' so others could replicate. The last two data collections took place over the course of an ecosystem unit that occurred in two portions (hence data collection IIIa and IIIb) spanning over three weeks. Students were challenged to research an animal of their team's choice to be the main character of a video game they would develop. Later, students designed their game by using a combination of manipulatives and mobile devices to create the components of their animal's ecosystem (e.g., predator and prey relationships, biotic and abiotic factors).

Measurement Tools

Participants from all three fourth-grade classrooms were surveyed pre- (beginning of the school year) and post- (end of the school year) participation in the learning experiences using the validated *Student Attitudes toward STEM (S-STEM) Survey* (FI, 2012) to evaluate changes in students' interest and attitudes towards STEM, careers, and workforce readiness skills. An adapted, digital version of the tool included a section reworded from science to computer science. Graphic cues can be helpful when collecting information from young children (Chambers & Johnston, 2002; Norman, 2012), so Emojis accompanied text for the survey's 4-point Likert scale. Related graphics were also used for STEM careers questions to scaffold completion of the survey. In order to keep consented participation transparent, all fourth-grade students participated in the survey, as well as through the additional methods of data collection.

Classroom observations were conducted using the *Engagement Observation Summary* (Activation Lab, 2016), a tool used to measure observed levels of student engagement. The observation data was used to verify information collected by additional survey and focus-group data, as it has been done in other research on engagement (Fredricks & McColskey, 2012). After each observation, students completed a self-report using the *Engagement in Science* tool (Activation Lab, 2016), which was followed by a focus group to help identify emerging themes regarding students' attitudes towards STEM+C, career interests, and workforce readiness skills. A tool to better organize the focus group data collection process was created by the researchers to facilitate the collection of data. Questioning strategies that presented themselves more 'game-based' were used during the focus group to accommodate for the younger age range. For example, Emoji signs were used to help students identify and describe their levels of affective and behavioural engagement, as well as their use of 4C skills from the activity. The data collection tool and game-like strategies provided a way to listen to what students had to say about their experiences and helped understand how the quality of learning experiences could affect student engagement (Dunleavy & Milton, 2009).

Validity

Validity was strengthened by using validated instruments (i.e., S-STEM, Engagement Observations Summary, Engagement in Science self-report, and EQUIP). Prior to data collection, the Emojis used for the S-STEM Likert scale were piloted with students of the same grade level as the study's participants. The EQUIP observation tool (Marshall et al., 2009) was used to measure the overall quality and variation of inquiry-based teaching observed (per Carlone et al., 2011). While the lessons were reviewed by experts before implementation, Carlone et al. (2011) used the EQUIP tool to rate the overall quality of implementation across different classrooms with different teachers as a proactive measure. Since this tool was only used as a guide to validate the quality of inquiry of an implemented lesson, only the tool's rubric for instructional and curriculum factors were used to evaluate lessons in this research study. The method of using these factors of the EQUIP tool to evaluate the level of inquiry of a lesson has been used in other studies (Henderson-Rosser et al., 2017).

Reliability

For reliability, there was transparency in the role of the observing researcher (Zohrabi, 2013), including documentation of time spent on the intervention. Utterances from the observations were used to code for the 4Cs, whereas utterances from the focus groups primarily used open codes. The triangulation of data across sources also provided a way to expose credibility to the case study findings exposing the coherent study design while maintaining the study's aim (Hyett et al., 2014). In addition, a statistical measure, Cronbach's alpha, was run to assess the internal consistency of the sets of scale and test items (Field, 2013) for both the S-STEM and the Engagement in Science self-report. The S-STEM reported high internal consistency (i.e., above 0.7) for all sets of test items. The Engagement in Science is self-report instrument, validated for overall engagement across three constructs with two sub-factors of the scale including an affective and combined cognitive/behavioral construct of engagement. However, Cronbach's alpha suggested that only the overall and affective test items had a high reliability (greater than 0.7), whereas the combined cognitive and behaviour constructs was low (i.e. 0.53) in the study administration and thusly removed from the final analysis.

Trustworthiness

To ensure trustworthiness, steps were taken to enhance qualitative data collection, analysis and interpretation. Researchers have criticized the reliability of using observation tools to measure engagement due to a lack of experience (Fredricks & McColskey, 2012), so five hours was dedicated to piloting the Engagement Observation Summary prior to the start of data collection. The observer found using field notes during observations was a more reliable method then recording directly on the protocol; such an adoption was suggested by the developers (Activation Lab, 2018). Further, this method provided an opportunity to neatly rerecord the field notes for the additional use in using the tool's coding procedure to quantify observed levels of engagement (i.e., overall, cognitive, affective/behavioural).

To ensure reliable coding, data collection I was first double coded by two researchers, mutually agreeing upon the NEA's (2012) definition of 4C constructs. Open codes were combined for similar meanings and the remaining data were analysed by the same two researchers. To measure intercoder reliability, percent agreement was calculated between coders on all four data collections. Percent agreement was 85%, 82%, 77%, and 80% respectively. A third researcher reconciled disagreements between coders one and two for each of the 4C constructs.

Limitations

Given the tailored nature of the intervention in the RPP collaboration, external validity is a limitation as different results may arise from use of the intervention in a different geographic setting with different students (Zohrabi, 2013). However, use of theory and best practices helps to mitigate these factors. Additionally, one of the researchers played the lead role in implementing most of the learning experiences with the students since the teachers were building their own confidence about implementing lessons with new resources through small-group work; however, the benefits to the teachers own experiences of integrating effective and engaging STEM+C out weighs the potental for limitation. Another limitation is the dearth of data collected to fully understand students' connections to STEM careers. Meaning, students interacted with STEM experts related to their learning content, but the researchers did not make any formal observations during these experiences. Lastly, observations of workforce readiness skills were not normally distributed amongst the three interventions, and creativity was observed the least although the interventions provided a lot of choice in design outcomes. However, creativity is a construct that is known to be difficult to observe reliably (Katz-Buonincontro & Anderson, 2018; Michael & Wright, 1989), so it was likely undercounted. Also, since the protocol only observed one student at a time, only their utterances were recorded and coded, possibly limiting the interpretation of the interactions with peers to the researchers coding that were not physically present during the activities. We acknowledge the common limitations (i.e., generalizability, reproducibility, and research bias) of case study research (Yin, 2018), however, the use of extant theory and validated protocols, coupled with the rich information the case study yielded suggests findings warrant a valuable contribution to the field of teaching STEM+C among geographically underrepresented (rural) elementary aged students.

Results

Analyses were conducted on data collected (approximately 10 hours per class) from each of the four sessions that modelled integrated STEM+C learning experiences that included 21 observations, 12 classroom sets of self-reports, and 11 focus groups. In addition, data was used from the pre- and post- S-STEM survey from 32 consented participants, as two students had only completed the pre-test and so it was not used in the final analysis. Quantitative S-STEM analyses were conducted per the author (Friday Institute for Educational Innovation [FI], 2012), employing t-tests for construct level (interval) data, Wilcoxon signed-rank tests for item-level (ordinal) data, and chi-square analyses for comparisons with categorial data collected. Qualitative analyses of workforce readiness skills in collaboration, critical thinking, communication, and creativity (i.e., the 4Cs) were based on the NEA (2012), *Preparing 21st Century Students for a Global Society: An Educator's Guide to the 'Four Cs'*, as *a priori* codes since they provide vetted and comprehensive definitions of each of these constructs, which have been used in other research on STEM education (Hite & McIntosh, 2020). The coding process produced frequencies of the 4Cs which were counted from activity observations. Frequency counts were documented during the focus groups, when the students were asked if they had interest in any of the twelve STEM careers presented to them graphically. (These same graphics were used to scaffold learning for the S-STEM instrument.) Since the focus groups generally had three to five consented students at a time, the sample providing frequency data was not large enough but often yielded documented feedback from students. Frequency counts did help record verbal reports from students for affective and behavioural constructs of engagement, used only for triangulating data from students' self-reports, that also suggested students' high levels of engagement affirmed by the observational data.

Quantitative Results

Descriptive statistics was first used to analyse the pre- and post-test data for attitudes by subject area constructs of the S-STEM survey and parsed by gender (Table 1). Next, a paired t-test was used to examine STEM attitudes by subject area constructs (i.e., aggregated, numeric data) between survey administrations finding a significant positive increase in females' attitudes from pre- (M = 3.97, SD = 0.38) to post-test (M = 4.35, SD = 0.30) in mathematics and pre- (M = 3.57, SD = 0.35) and post-test (M = 4.05, SD = 0.30) attitudes in science. When the data from males and females were bounded, science attitudes from pre- (M = 3.49, SD = 0.42) to post-test (M = 3.94, SD = 0.32) evidenced a significant increase.

Table 1
Attitudes by Construct Averages Between S-STEM Survey Administrations

Construct	Number of Items	Pre-Administration of S-STEM Average			Post-Administration of S-STEM Average		
Attitudes about:		Total (N=32)	Females (N=19)	Males (N=13)	Total (N=32)	Females (N=19)	Males (N=13)
Science	9	3.49	3.57	3.37	3.94	4.05	3.77
Mathematics	8	4.10	3.97	4.29	4.40	4.35	4.47
Computer Science	9	3.34	3.32	3.38	3.68	3.75	3.57
Technology and Engineering	9	3.99	3.92	4.09	4.11	4.09	4.13
21st Century Learning	11	4.35	4.53	4.09	4.39	4.54	4.18

Note. Responses based on a 5-point Likert scale, Strongly Agree (5) to Strongly Disagree (1).

The Wilcoxon signed-rank test was employed to compare pre- and post-test data, at the item-level (i.e., ordinal, Likert data), from the S-STEM survey. Four question items showed positive significance including choosing a career in science from pre- (Mdn = 3.00) to post-test (Mdn = 3.00), T = 176, p = .031, r = 0.38, knowing science will help earn money when they are older pre- (Mdn = 3.00) to post-test (Mdn = 4.00), T = 202, p = .013, r =0.44, needing to understand science for a job when they are older pre- (Mdn = 4.00) to post-test (Mdn = 4.00), T = 157, p = .012, r =0.45, and thinking computer science is not so hard to understand pre- (Mdn = 3.00) to post-test (Mdn = 4.00), T =236, p = .045, r = 0.35.

S-STEM results further suggested that students' perceptions increased from pre- to post-test of their mathematics (75% to 84%) and science (59% to 75%) performance. Additionally, there were gains in their knowledge of STEM+C careers with the largest gains in girls' knowledge of scientists and computer scientists (Table 2). An expanded section of the survey asked students to identify sources of information in which students' gained knowledge of STEM+C professionals including school and/or textbooks, magazines and/or books, home and/or family, television, and internet or social media. Students reported gains in knowing STEM+C professionals from school and/or textbooks and internet or social media. For example, at the beginning of the school year students reported 6% of their knowledge of computer scientists came from school and 16% from internet related sources. At the end of the year, 41% came from school and 37% came from the internet. Similar gains in both these sources of information were found in all subject domains, except mathematics. The post-test data showed only 9% of students knew of a mathematics career from school and 28% from the internet.

Table 2
Knowledge of STEM Professionals by Averages Between S-STEM Survey Administrations

Type of STEM Career	Pre-Administration of S-STEM			Post-Administration of S-STEM		
	Total (*N*=32)	Females (*N*=19)	Males (*N*=13)	Total (*N*=32)	Females (*N*=19)	Males (*N*=13)
Scientists	17 (53%)	9	8	26 (81%)	16	10
Engineers	21 (66%)	9	12	26 (81%)	15	11
Mathematicians	13 (41%)	7	6	16 (50%)	11	5
Technologists	18 (56%)	10	8	25 (78%)	16	9
Computer Scientists	16 (50%)	7	9	26 (81%)	16	10

Note. Responses were a Binary choice, Yes (1) or No (0), which represents who they knew.

After each activity that was observed, the Engagement in Science self-report was given to the students. The survey data was analysed with descriptive statistics and found that the students reported very high to high levels of affective and overall engagement (i.e., affective, cognitive/behavioral) during the four data collections (Table 3). Notably, the lessons across all the classrooms were evaluated for levels of inquiry using the EQUIP tool, and results suggested that all the activities were proficient or exemplar models of inquiry-based learning.

Table 3
Students' Self-Reports of Engagement and Affect, including Overall Averages

Activity	N-Size	Affective Construct Score	SD	Overall Engagement Score	SD
Data Collection I	31	1.68	0.76	1.65	0.85
Data Collection II	29	1.45	0.56	1.49	0.68
Data Collection IIIa	32	1.26	0.42	1.38	0.74
Data Collection IIIb	24	1.24	0.46	1.36	0.73

Note. Responses based on a 4-point Likert Scale, YES! (1) to NO! (4).
The self-report is validated to make inferences from two sub-factors of the scale (i.e., affective score or a behavioral/cognitive score), but a low reliability for a behavioral/cognitive score merited it being eliminated from the findings. The survey is also validated to make inferences regarding the overall engagement (i.e., a combination of affective, behavioral, and cognitive engagement).

Qualitative Results

The observation protocol, supplemented by researcher field notes, provided summative values on constructs related to engagement. In addition, a combination of the observation and focus group utterances were coded by the researchers for workforce readiness skills (i.e., the 4Cs) and open codes were developed and merged for further analysis. The open codes documented utterances that illustrated a transfer of knowledge/connection, personal interest and/or ability, but were primarily found in the coding of the focus group data. Workforce readiness skills were analysed by using the frequency counts of utterances coded for the 4Cs for each of the four activities. In each section, the specific skill is denoted in braces for the reader. Chi-square analyses of independence were run to examine relationships between the activities and the frequencies of observed skills. Significant differences were found, X^2 (1, N=280) = 21.85, p = .009, meaning the skills were not observed consistently across the four activities. Data collection I had the most observed actions related to students observed use of the 4Cs (N = 88), followed by IIIb (n = 72), II (n = 68), and IIIa (n = 52). Data collection I had more observations of critical thinking (n = 41) and collaboration (n = 22) observed as compared to communication (n = 19) and creativity (n = 6). For example, Sarah [pseudonym] was observed critically thinking through collaboration with her group as, "well, then we have to start her [robot] and go through here," referring to how to plan the robot's path through the maze. Students in data collection I were tasked with ways to think critically; Jane described their skill use in the focus group as:

> we were going to discuss with the group {collaboration} and then the car was going bad and now we had to like scratch, um, what we were first working on. We're working on, we have to sketch it because we have to go across the road and not under it {critical thinking}.

Data collection II, had more observations of critical thinking (n = 29) and communication (n = 23), while collaboration (n = 11) and creativity (n = 5) were less observed. Lee was determining the type of circuit by thinking aloud in his group {communication} that it was "parallel because if we take a light out (takes light out) {critical thinking}, it still works!" In this activity communication was key, Julie admitted in the focus group when communication broke down, "…it was hard working with somebody that was trying to kind of do it all…I tried to add designs, but she, um, just covered them with playdough." Data collection IIIa, had more observations of collaboration (n = 19) and critical thinking (n = 13) and less of communication (n = 11) and creativity (n = 9). Students actively collaborated in the co-construction process, Sean exclaimed, "we are making the guppy's home, it's a coral reef!" This sentiment was amplified by Carol in the focus group when they said, "Yeah, we all worked in the group to make the character. We all figured out things together {critical thinking} and we didn't leave anybody out {collaboration}." Data collection IIIb, had more observations of critical thinking (n = 26) and communication (n = 22) and similar observations of collaboration (n = 11) and creativity (n = 13). Interestingly, by this final data collection point, students were more able to communicate their ideas and identify problems productively. Joyce during the activity identified that, "we need to add the seaweed," and Heather raised their concerns in stating, "that's not white, it's orange." Improved communication provided avenues for troubleshooting, as Morgan shared in the focus group, "when we figured out Cactus Boy could not be our main character and go back and make it snake and re-plan {critical thinking}." Altogether, reports of critical thinking were most observed (n = 109), followed by communication (n = 75), collaboration (n = 63) and creativity (n = 33).

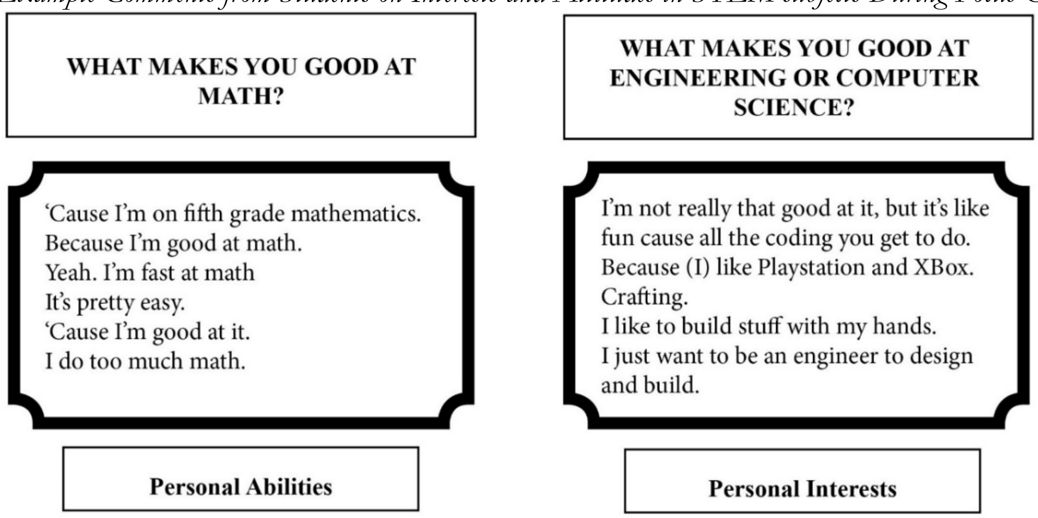
Case Analysis and Discussion

Overall, the study aimed to understand how integrated STEM+C lessons, developed through a RPP, could augment fourth grade students' attitudes and interests towards STEM+C. Case results suggest significant positive gains in interests and attitudes in science for all students, with the greatest (significant) benefit for girls in both science and mathematics. This finding supports a study by Grover et al. (2014) who found students' experiences in CS helped them to not only understand CS, but also develop their appreciation for its applicability across disciplinary domains. Findings also support existing research that suggests integrative STEM learning experiences can positively support students' attitudes (Toma & Greca, 2017; Tran, 2018) and interests (Lamb et al., 2015) in specific areas within STEM.

Further, students reported improved perceptions of their abilities to perform in science and mathematics. Focus group data suggested that students often equated their personal ability for math to their high performance in mathematics, however, when it came to engineering or CS, their perceptions relied more heavily on personal interests than innate ability (see Figure 2). Comments pertaining to the other subject areas were similar, meaning students balanced perceptions of their abilities in the subject with their interest in the subject. These findings extend current thinking to how students' evaluations through formal assessment (e.g., grades) become conceptualized by students as the ability to learning facts and formulas, as opposed to their perceptions of inherent ability (Haimovitz & Dweck, 2017). This study indicates this ability-interest negotiation occurs at very young ages and is important considering perceptions of ability helps to forge one's personal interest (in science). Further, these findings support the need for educators to endorse a deeper learning process within science as well as other subject areas.

Figure 2

Example Comments from Students on Interests and Attitudes in STEM subjects During Focus Groups



Much of the S-STEM data for STEM career aspirations had a ceiling effect from pre- to post-administrations; however, student awareness increased in the different STEM-based careers. The greatest increase was in knowledge of professionals in STEM+C careers, especially for sampled girls. Meaning, knowledge of professionals in CS doubled from pre- to post-intervention for girls in every career except mathematics. Students reported they garnered their STEM career knowledge largely through school and/or textbooks and the internet or social media, suggesting the students were 'seeing' more STEM career experts both in- and out-of-school. Greater exposure to STEM careers affirms related research that also found career knowledge growth among middle-grades students when watching STEM career videos (Kier et al., 2014). Findings suggest such experiences to generate career awareness are beneficial for elementary learners, too.

Workforce ready skills, categorized by the 4Cs, were observed throughout the intervention, but were not equally among the four activities. Communication and collaboration were most evidenced with creativity as the lowest observed construct across all of the activities. First, there were over 30 more observed actions between data collection I and IIIa. Data collection I, a challenge that involved using probeware to measure friction in conjunction with learning a new tool to code and help simulate the outcomes of a designed prototype had the greatest number of observations that included critical thinking and collaboration, suggesting these skills were used more frequently to complete the inquiry-based learning experience that incorporated many new learning experiences merging together into one learning experience. Data collection II, a paired student activity where students got to create their own object out of conductive playdough, had more observations of critical thinking and communication than expected, suggesting that decreasing the number of students to collaborate can increase the communication in the group to problem-solve. These specific findings affirm current research that shows workforce readiness skills can be supported through inquiry-based learning activities (Tran, 2018). Further, findings suggest consideration of cooperative groupings of students (e.g., size) is vital to maximize use of those skills.

Data collection IIIa and IIIb also had interesting results, as they varied in observations across the 4Cs although the students were engaged in the same learning project. In the third learning unit of the intervention, students researched and developed a plan to create a video game to illustrate an animal's ecosystem. In both data collections for their project, students were actively building their video games through the use of manipulatives to create their main character and the interactions the organism had with its ecosystem through adding assets and animations to their game scenes to include items like sources of food, shelter, and water. However, data collection IIIa observations suggested collaboration was observed the most and for data collection IIIb, critical thinking was observed the most, which may have to do with the idea that while the students were actively iterating their video game designs, more troubleshooting (e.g., debugging program glitches) occurred to make the products play better for the audience as time progressed in the project.

Creativity lacked the most in overall observed occurrences across all of the learning activities. While all of the activities provided constraints, they also provided student choice in how they wanted to creatively illustrate their learning outcomes. For instance, data collection II had the least number of occurrences observed and coded for creativity, yet the activity gave students the least amount of constraints to create a product. Data collection IIIa and IIIb had the most occurrences of creativity, suggesting the activity fostered an environment for the students to be creative, but the overall lower frequency of observed creativity across all the learning experiences suggest that it was harder to observe (Katz-Buonincontro & Anderson, 2018). Regardless, this suggests teachers should leverage explicit avenues for students to engage in creativity activities, so they can be used, grown, and observed.

In regard to engagement, throughout the intervention students reported high to very high levels of engagement in their self-reporting and focus group data (see Table 3). The affective engagement and overall engagement gradually increased with every activity observed, including the last two activities which had the same learning objectives, but took a longer period of time to implement. Fewer observations initially may be attributed to the inexperience of sampled students with small group, inquiry-based learning, which grew over time. As the intervention progressed, students began to engage more in the activities as they got accustomed to active participation, extending current research suggesting that inquiry-based learning activities need to also be engaging to positively support STEM learning opportunities (DeJarnette, 2012; Finn & Zimmer, 2012; Trowler, 2010). This indicates that younger students will need scaffolding for STEM+C activities that task them with employing 4C-based skills.

Conclusion

Results suggest among the 34 underrepresented, rural fourth students that had participated in the three RPP designed inquiry-based learning modules integrating STEM+C, sampled students' interests, attitudes and workforce readiness skills (within the 4Cs) had improved. The intervention also proved to be positively engaging to students. One of the largest impacts appear to be how the students were able to connect their context of learning and relate it to STEM career opportunities. While this study provides an opportunity to examine an intervention over a year-long period, it still leaves a gap in the number of longitudinal studies that measure student engagement (Reschly & Christenson, 2012). Further research needs to be designed to measure changes in students' interests and attitudes in STEM+C and STEM+C careers from elementary to secondary education, where students begin to choose courses for their career pathways, after participating in STEM+C interventions.

The authors received no financial support for the research, authorship, and/or publication of this manuscript.

Stephanie D. C. Playton (playtonsd@longwood.edu) is a STEM Learning Specialist at the College of Education, Heath, and Human Services at Longwood University in Farmville, Virginia. Her current role supports STEM education for PK-12 students and teachers in under-served, rural schools. Current research interests include integrated STEM, STEM awareness and interest, and computational thinking. She is also a PhD candidate with Texas Tech University in Lubbock, Texas.

Rebecca Hite (rebecca.hite@ttu.edu) is an assistant professor of STEM education in the Department of Curriculum and Instruction at Texas Tech University, in Lubbock Texas. Her research foci explore cognitive and non-cognitive affordances of emergent technologies in K-12 STEM instruction, focusing on under-presented and under-served groups in STEM.

Paula K. Leach (leachpk@longwood.edu) is the director of the Institute for Teaching through Technology and Innovation Practices at Longwood University in Farmville, VA: an organization focused on PK-12 outreach to rural school districts in the south central part of Virginia. She has been an educator for over 25 years and enjoys working with teachers to introduce them to technologies and strategies that provide engaging opportunities for students. She received her undergraduate degree from Slippery Rock University in Slippery Rock, PA and graduate degrees from Virginia Commonwealth University in Richmond, VA.

References

2018 State of Computer Science Education. (2018). *2018 State of Computer Science Education.* https://code.org/files/2018_state_of_cs.pdf

Activation Lab. (2016). *Measures Technical Brief: Engagement in Science Learning Activities, version 3.2.* http://activationlab.org/wp-content/uploads/2018/03/Engagement-Report-3.2-20160803.pdf

Activation Lab. (2018). *The Activation lab evaluation toolkit.* http://activationlab.org/toolkit/#UsingYourResources

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of Research on Student Engagement* (pp. 283–302). New York, NY: Springer Science.

Ball, C., Huang, K. T., Cotten, S. R., & Rikard, R. V. (2017). Pressurizing the STEM pipeline: An expectancy-value theory analysis of youths' STEM attitudes. *Journal of Science Education and Technology*, *26*(4), 372–382.

Baker, R. (2003). Promoting a research agenda in education within and beyond policy and practice. *Educational Research for Policy and Practice*, *2*(3), 171–182.

Barzanji, T. (2013). *Growing the STEM: Encouraging Interest in STEM subjects among low socio-economic Australian secondary students.* Cambridge, MA: Harvard Kennedy School.

Biriescu, S., & Babaita, C. (2014). Rural education, an important factor of regional development in the context of local government strategies. *Procedia - Social and Behavioral Sciences*, 77–86. https://doi.org/10.1016/j.sbspro.2014.02.462

Bureau of Labor Statistics, U.S. Department of Labor. (n.d.). Occupational Employment Statistics. https://www.bls.gov/oes/topics.htm

Carlone, H. B., Haun-Frank, J., & Webb, A. (2011). Assessing equity beyond knowledge- and skills-based outcomes: A comparative ethnography of two fourth-grade reform-based science classrooms. *Journal of Research in Science Teaching*, *48*(5), 459–485.

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, *27*(1), 27–36.

Christensen, R., Knezek, G., & Tyler-Wood, T. (2015). Alignment of hands-on STEM engagement activities with positive STEM dispositions in secondary school students. *Journal of Science Education and Technology*, *24*(6), 898-909.

Christenson, S. L., Wylie, C., & Reschly, A. L. (Eds.). (2012). *Handbook of research on student engagement.* Springer Science.

Code.org Advocacy Coalition. (2018). *2018 Computer science education: Policy and implementation.* https://advocacy.code.org/2018_state_of_cs.pdf

DeJarnette, N. K. (2012). America's children: Providing early exposure to STEM (science, technology, engineering and math) initiatives. *Education*, *133*(1), 77–84.

Dorph, R., Bathgate, M. E., Schunn, C. D., & Cannady, M. A. (2018). When I grow up: the relationship of science learning activation to STEM career preferences. *International Journal of Science Education, 40*(9), 1034-1057.

Dunleavy, J., & Milton, P. (2009). *What did you do in school today? Exploring the concept of student engagement and its implications for teaching and learning in Canada*. Canadian Education Association.

Eccles, J. & Wang, M. (2012). So what is student engagement anyway? In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of research on student engagement* (pp. 133–148). Springer Science.

Ehren, B. J., & Murza, K. (2010). The urgent need to address workforce readiness in adolescent literacy intervention. *Perspectives on Language Learning and Education*, *17*(3), 93-99.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

Finn, J.D. & Zimmer, K.S. (2012). *Student engagement: What is it? Why does it matter?* In S.L. Christenson, A.L, Reschley, & C. Wyle (Eds.), *Handbook of research on student engagement* (p. 97-131). Spring Science + Media.

Fredricks, J. A. & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of research on student engagement* (pp. 763–782). New York, NY: Springer Science.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507–537.

Friday Institute for Educational Innovation. (2012). *Student Attitudes toward STEM Survey – Upper Elementary School Students,* Author.

George, R. (2000). Measuring change in students' attitudes toward science over time: An application of latent variable growth modeling. *Journal of science Education and Technology*, *9*(3), 213-225.

Grover, S., Pea, R., & Cooper, S. (2014, March). Remedying misperceptions of computer science among middle school students. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 343-348).

Guzey, S. S., Moore, T. J., Harwell, M., & Moreno, M. (2016). STEM integration in middle school life science: Student learning and attitudes. *Journal of Science Education and Technology*, *25*(4), 550–560.

Haddock, G., & Maio, G. R. (Eds.). (2004). *Contemporary perspectives on the psychology of attitudes : An introduction and overview*. https://ebookcentral.proquest.com

Haimovitz, K., & Dweck, C. S. (2017). The origins of children's growth and fixed mindsets: New research and a new proposal. *Child Development, 88*(6), 1849–1859.

Heil, D., Pearson, G., & Burger, S. (2013, June 23-26). *Understanding integrated STEM education: Report on a national study*. Paper presented at *ASEE Annual Conference*. Atlanta, GA.

Henderson-Rosser, A., & Sauers, N. J. (2017). Analyzing the effects of one-to-one learning on inquiry-based instruction. *Computers in the Schools*, *34*(1–2), 107–123.

Hite, R., & McIntosh, A. (2020). The Affordances of 3D Mixed Reality in Cultivating Secondary Students' Non-cognitive Skill Use and Development in the Engineering Design Process. In R. Zheng (Ed.), *Cognitive and affective perspectives on immersive technology in education* (2nd ed., pp. 171-194). Hershey, PA: IGI Global. DOI: 10.4018/978-1-7998-3250-8.ch009

Hyett, N., Kenny, A., & Dickson-Swift, V. (2014). Methodology or method? A critical review of qualitative case study reports. *International Journal of Qualitative Studies on Health and Well-Being*, *9*(23606), 1-12.

Johnson, J., & Strange, M. (2007). *Why rural matters 2007: The realities of rural education growth. Rural School and Community Trust*. http://eric.ed.gov/?id=ED498859

Katz-Buonincontro, J., & Anderson, R. C. (2018). How do we get from good to great? The need for better observation studies of creativity in education. *Frontiers in Psychology*, *9*(NOV), 1–5.

Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The development of the STEM career interest survey (STEM-CIS). *Research in Science Education*, *44*(3), 461-481.

Kotok, S., & Kryst, E. L. (2017). Digital technology: A double-edged sword for a school principal in rural Pennsylvania. *Journal of Cases in Educational Leadership*, *20*(4), 3–16.

Kurz, M. E., Yoder, S. E., & Zu, L. (2015). Effects of exposure on attitudes towards STEM interests. *Education*, *136*(2), 229–241.

Lam, S., Wong, B., Yang, H., & Liu, Y. (2012). Understanding student engagement with a contextual model. In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of research on student engagement* (pp. 403–420). Springer Science.

Lamb, R., Akmal, T., & Petrie, K. (2015). Development of a cognition-priming model describing learning in a STEM classroom. *Journal of Research in Science Teaching*, *52*(3), 410–437.

Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science achievement. *Science Education*, *97*(4), 574-593.

LeMahieu, P. G., Nordstrum, L. E., & Potvin, A. S. (2017). Design-based implementation research. *Quality Assurance in Education*, *25*(1), 26–42.

Li, Q., Moorman, L., & Dyjur, P. (2010). Inquiry-based learning and e-mentoring via videoconference: A study of mathematics and science learning of Canadian rural students. *Educational Technology Research and Development*, *58*(6), 729–753. https://doi.org/10.1007/s11423-010-9156-3

littleBits. (2018). *Early exposure to STEM and its impact on the future of work* [White Paper]. https://e.littlebits.com/download-early-exposure-stem

Marshall, J. C., Smart, J., & Horton, R. M. (2009). The design and validation of EQUIP: An instrument to assess inquiry-based instruction. *International Journal of Science and Mathematics Education*, *8*(2), 299–321.

Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In J. A. Glober, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 33-52). Springer.

Monk, D. H. (2007). Recruiting and retaining high-quality teachers in rural areas. *The Future of Children*, *17*(1), 155–174.

National Education Association. (2012). *Preparing 21st Century Students for a Global Society.* http://www.nea.org/assets/docs/A-Guide-to-Four-Cs.pdf

National Science Board. (2015). *Revisiting the STEM workforce - a companion to science and engineering indicators 2014*. National Science Foundation.

National Science Foundation. (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017 digest.* Author.

Niu, L. (2017). Family socioeconomic status and choice of STEM major in college: An analysis of a national sample. *College Student Journal*, *51*(2), 218–312.

Noonan, R. (2017). *STEM Jobs: 2017 Update (Issue Brief # 02-17).* https://files.eric.ed.gov/fulltext/ED594354.pdf.

Norman, R. R. (2012). Reading the graphics: What is the relationship between graphical reading processes and student comprehension? *Reading and Writing*, *25*(3), 739–774.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Partnership for 21st Century Learning. (2015). *P21 framework definitions.* http://static.battelleforkids.org/documents/p21/P21_Framework_Definitions_New_Logo_2015_9pgs.pdf

Reschly, A., & Christenson S. (2012). Jingle, jangle and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, C. Wylie, & A. L. Reschly (Eds.), *Handbook of research on student engagement* (pp. 3–20). Springer Science.

Sarama, J., Clements, D., Nielsen, N., Blanton, M., Romance, N., Hoover, M., … McCulloch, C. (2018). *Considerations for STEM education from preK through grade 3*. Education Development Center, Inc.

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science education*, *96*(3), 411-427.

Skamp, K. (2007). Conceptual learning in the primary and middle years: The interplay of heads, hearts and hands-on science. *Teaching Science*, *53*(3), 1-5.

Sawchuk, S. (2017). *Virginia mandates computer science learning.* http://blogs.edweek.org/edweek/curriculum/2017/11/virginia_mandates_computer_science_learning.html

The New Teacher Project. (2018). *The opportunity myth: What students can show us about how school is letting them down - and how to fix it.* https://opportunitymyth.tntp.org/

Toma, R. B., & Greca, I. M. (2018). The effect of integrative STEM instruction on elementary students' attitudes toward science. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(4), 1383–1395.

Tran, Y. (2018). Computer programming effects in elementary: Perceptions and career aspirations in STEM. *Technology, Knowledge and Learning*, *23*(2), 273–299.

Trowler, V. (2010). Student engagement literature review. *Higher Education*, (November), 1–74. DOI: 10.1037/0022-0663.85.4.571

Webb, M., Davis, N., Bell, T., Katz, Y., Reynolds, N., Chambers, D. P., & Sysło, M. M. (2017). Computer science in K-12 school curricula of the 2lst century: Why, what and when? *Education and Information Technologies*, *22*(2), 445–468.

Xie, Y., Fang, M., & Shauman, K. (2016). STEM Education. *Annual Review of Sociology*, *41*(2), 331–357.

Yin, R. K. (2018). *Case Study Research and Applications* (6th ed.). Sage Publications.

Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies*, *3*(2), 254–262.

ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 25, NO. 2, 47-67

The Effects of the Flipped Classroom Model on the Laboratory Self-Efficacy and Attitude of Higher Education Students

Harun Çelik
University of Kırıkkale

Hüseyin Miraç Pektaş
University of Kırıkkale

Orhan Karamustafaoğlu
University of Amasya

ABSTRACT

In class environments, the flipped classroom (FC) model has been found to increase students' attitudes, achievement, and motivation. However, the practical effects of the FC model in laboratory environments have not been introduced into the literature. Therefore, in this study, the effects of the FC model on the laboratory self-efficacy skills and attitudes of higher education students were investigated within the scope of the physics laboratory course. The data were purposively collected from a group of 84 first-year university students aged between 18-20, who were, then, sorted out into two groups: experimental and control. The sequential explanatory design model by Creswell was used, which is a subcategory of the mixed-methods design. While the FC model was applied to the experimental group, the traditional classroom model was used in the control group. In the data collection process, qualitative and quantitative data collection tools were used sequentially. The experimental results obtained at the end of a six-week study showed that the FC model significantly improved the laboratory self-efficacy and attitude towards the laboratory of higher education students. Therefore, the FC model was also found to have positive effects in laboratory environments.

Keywords: flipped classroom model, augmented reality, information technologies, physics laboratory, science education

Introduction and Theoretical Framework

The common point of the solutions for maximizing student-teacher interactions is to make students more active by teaching them how to learn in a student-centred environment (Brewer & Movahedazarhouligh, 2018). The targeted development in the learning environment can be achieved by any method or pedagogical approach in which students are active. Active learning expresses how students participate in activities, reflect their ideas and use them. Active learning is a broad-based term consisting of different teaching methods to involve students in their learning and to develop and maintain a higher level of learning (Zepke, 2013). Therefore, active learning has made a popular teaching approach in higher education. Many studies have shown that the use of active learning methods can increase learning outcomes without compromising on the content (Owens, Sadler, Barlow, & Smith-Walters, 2017). Educators who understand the importance of active learning have developed new strategies to activate students in the learning process (Findlay-Thompson &

Mombourquette, 2014). However, the strategies developed within the scope of active learning limit the amount of time students need to actively use in the classroom. Teachers may not practise enough in their learning environment. Because they devote most of their time not to practising but to teaching theoretical knowledge, students may not be able to communicate with their peers or teachers (King & Newmann, 2001). According to some research, the FC model provides a solution to this problem. Thus, recently, educators have supported the FC model, claiming that students have more opportunities in active learning as this model promotes participation in interactive and high-level activities (Chuang, Weng, & Chen, 2016). The FC model is popular instructional model, in which activities traditionally conducted in the classroom become home activities, and activities normally constituting homework become classroom activities (Bergmann & Sams, 2012; Sohrabi & Iraj, 2016). The FC model based on active learning (Chen, Wang, & Chen, 2014) has recently become one of the most popular technology-based learning models (Jensen, Kummer, & Godoy, 2015). The FC model is seen as an important tool that provides opportunities for students in higher education to learn on their own. In addition, the FC model is widely accepted by educators because it defines the needs and capacities of individuals, offers individualized teaching and provides flexibility on planning the learning, gives homework and adjusts the pace of learning (Davies, Dean, & Ball, 2013). According to all these studies, it can be said that the FC model supports the learning environments as a process that enhances the learning experiences and encourages students to take responsibility in order to develop their learning products.

According to Awidi and Paynter (2019), the FC model can provide cooperation between instructional technologies that support appropriate student behavior and learning, and provide a driving force for students to actively participate in their learning. By encouraging discussion and collaboration, the FC model requires students to learn concepts and ideas through an in-depth study of content in contrast to passive superficial learning (Burke & Fedorek, 2017). Therefore, the use of the FC model in K12 education and higher education has been increasing (Álvarez, 2012). The FC model is defined as an innovative approach in teaching, which has the potential to create active, busy and learning-centred classes (Brewer & Movahedazarhouligh, 2018). The FC model is also called inverted instruction or inverted learning, but these terms refer to the same teaching strategy (Fidalgo-Blanco, Martinez-Nuñez, Borrás-Gene, & Sanchez-Medina, 2017). This model is a special learning type that requires active participation of students in learning activities before and during face-to-face lessons with teachers (Strayer, 2012). At the same time, in the learning environment, this model is a solid pedagogical technique that improves students' comprehension appreciation of the lesson (Samuel, 2019).

FC Model Benefits and Limitations

There are many benefits resulting from the use of the FC model. It is suitable for all learners because it has positive effects and can focus on all the students (Sams & Bergmann, 2013). Unlike a teacher-centred teaching model based on traditional learning, the FC model has two inverted education stages (Bergmann & Sams, 2012). The first part of the FC model is the planning phase which is the stage before the course. At this stage, students acquire the basic conceptual information with online materials (Bergmann & Sams, 2012; Strayer, 2012), so they often receive most of the information transfer before attending the classroom (Abeysekera & Dawson, 2015). The second part of the FC model is the in-class learning phase. At this stage, there are active learning activities in the classroom environment such as laboratory experiments, interactive lessons and problem solving (Strayer, 2012). Class time is used for active and social learning activities in this method (Abeysekera & Dawson, 2015). In other words, the content presented in the classroom traditionally such as face-to-face learning is given to the student prior to the class as homework (Herreid & Schiller, 2013), and students can access the lessons at their own pace (Moffett, 2015). However, in order to increase student motivation for courses with active learning activities, students should attend face-to-face courses at the preliminary information and preparation level (Hao, 2016) because teachers play a vital role in students' lives and having face-to-face interaction with teachers is an invaluable experience for them (Bergmann & Sams, 2012). Out-of-class learning is flexible in the FC model. Students can match their academic level and individual needs according to their preference, and this matching can take place at any time and in any place (Moffett, 2015). In the FC model, there remains more time for students to apply what they learn, to develop high-level thinking skills, to make classroom discussions, to focus on projects and problem solving (Hwang, Lai, & Wang, 2015). Therefore, this model increases the time spent in the classroom for active learning. Research that takes into account all these situations shows that the FC model positively affects students' learning (Davies et al., 2013; Er, Kopcha, Orey, & Dustman, 2015; Schultz, Duffield, Rasmussen, & Wageman, 2014; Shih, & Tsai, 2017; Strayer, 2012). FC practices show that students' satisfaction and motivation may lead to differentiation or decrease. The basis of this perception is the possibility of decreasing the frequency and willingness of participation in the activities as opposed to the expectations of the students. Therefore, the researchers stated that students were concerned about the resistance to participation in the activities and that this could reduce the efficiency of the FC model (Gençer, 2015; Hao & Lee, 2016; Herreid & Schiller, 2013; Missildine, Fountain, Summers & Gosselin, 2013; Wilson, 2013). If the FC model is used in all courses, this may make it difficult for students to prepare adequately for the courses (Hao, 2016; Wanner & Palmer, 2015). Students may not have the necessary digital skills to manage a technology-integrated environment (Hao & Lee, 2016).

Self-Efficacy and Attitude in Technology-supported Science Laboratories

The rapid expansion of technology in schools and its integration into teaching have caused the ways for teacher, student and curriculum interaction to be reconstructed. As a result, changes have occurred in learning environments. The constructivist learning environment which adopts student-centred active learning methods is also integrated with computer-assisted teaching methods (Çelik & Pektaş, 2017). Among these methods, Augmented Reality (AR) applications are the new and popular applications. Research shows that AR applications have gained popularity because of being flexible, easy-to-use, user-friendly and low-cost when supported by mobile learning (Jones & Jo, 2004). Wearable technologies can be used to integrate multimedia tools such as video, audio and graphics into learning environments (Churchill & Hedberg, 2008). As a result of its integration into written materials, learners can establish a connection between these tools; thus, learning can result in meaningful and profound (Burden & Kearney, 2016). The applications which are evaluated within the methods known as mobile learning are evaluated as the applications of Quick Response (QR) code in education (Chen, Chang, & Wang, 2008). The environments where cooperative learning is best practiced are experimental study environments in science and technology laboratories.

Combining technology with physics laboratories provides a different experience for students. This experience is an opportunity to provide an alternative to difficult-to-reach, expensive, dangerous and complex experiments (Akçayır, Akçayır, Pektaş, & Ocak, 2016). Computer technology supported by Internet applications and multimedia is used in educational activities to increase learners' motivation and learning towards physics. Students prefer to use web pages, interactive video features and recorded videos outside the classroom to support learning due to technological advances, while in the classroom they prefer to take responsibility in active learning activities such as discussion, problem solving and group work. Hence, it can be said that FC practices are the best models that can be used in laboratory courses, which can move the student's attitude with video and various animations to the upper level, and enable them to develop laboratory self-efficacy. Self-efficacy is defined as people's beliefs in their ability to achieve a certain level of performance (Bandura, 1994). Teaching strategies focusing on providing students with opportunities for performance are well aligned with the emphasis on active achievement (actual performance gained through direct experience) as the most effective source of the knowledge of self-efficacy (Bandura, 1977). The self-efficacy component of the social-cognitive theory by Bandura is critical for academic achievement, motivation and learning, and often reviewed in academic studies (Pajares, 1996). The majority of the studies supporting this view found positive relationships between self-efficacy and performance and attitudes (Sitzmann & Yeo, 2013). This relationship can be explained by the effect of having a high level of self-efficacy belief as well as many other factors in students' performance. Therefore, it is important to consider the necessity of high self-efficacy beliefs as well as a positive attitude towards active learning in the laboratory environment in order for the students to perform laboratory activities effectively and successfully.

Purpose Statement

Studies in the literature focusing on using the FC model in physics teaching (Aşıksoy & Özdamlı, 2016; Akı & Gürel, 2017; Deslauriers, Schelew, & Wieman, 2011; Gómez-Tejedor et al., 2020) are very few. In addition, the laboratory self-efficacy and learning outcomes of students were largely ignored in these studies. Therefore, the integration of both the FC model and the learning outcomes in the laboratory will contribute significantly to the science education literature. The aim of this study is to determine whether the use of the TC (Traditional Classroom) model or the FC model has a significant difference on the laboratory self-efficacy of the students. To this end, the study of the subject of electricity is important in terms of its contribution to the literature because according to the literature, one of the issues that are difficult to be conceptualized and have the potential to be misunderstood is the subject of electricity in the general physics course (Akbaş & Pektaş, 2011). Studies have revealed that students face problems in understanding electricity, they have a risk of misconceptions and they have difficulty in analyzing abstract problems (Chambers & Andre, 1997; Sencar & Eryilmaz, 2004). Even adults admit that electricity is difficult to understand, similarly to other subjects in physics (Shipstone, 1998).

Physics laboratories are considered as learning environments where students make high-level conceptual learning (Çepni, Kaya, & Küçük, 2005). In the realization of this conceptual learning, attitude is in the foreground. Increasing student achievement in laboratory courses shows a relationship with student attitudes towards physics laboratory (Palic & Pirasa, 2012). Attitude is also a positive factor for students' academic achievement in science (Osborne, Simon, & Collins, 2003). In physics laboratories, it is important to investigate the effects of the use of new models such as FC on student attitudes and for the contribution of FC to the literature as well. So, the second aim of this study was to determine whether the use of the TC model with the FC model had a significant difference on the attitudes of the students. With this study, it is thought that a study into the positive and negative effects of FC will contribute to the literature through the opinions and suggestions of the students in the experimental group. The final aim of the study is to describe the practical views of the students in the experimental group in which the FC model is applied. Based on the stated purposes, the research questions (RQs) were determined as follows: *RQ #1*: Is there a significant difference in terms of the laboratory self-efficacy between the experimental and control group? *RQ #2*: Is there a significant difference in attitude scores between the experimental and control groups? *RQ #3*: What is the opinion of the experimental group for the FC model?

Methodology

This study employs the sequential explanatory design model under Creswell's mixed models, which is suitable for both qualitative and quantitative data analysis. On the other hand, qualitative data was used to increase, expand or support quantitative data. An analysis of the data is conducted in an interrelated manner and is often combined in the data interpretation and discussion (Creswell, 2003). The internal validity of the qualitative dimension of the research was ensured by variations and participant confirmation. The external validity was attained by making direct quotations from the basic features of a descriptive analysis and by interpreting the data in detail (Özmen & Karamustafaoğlu, 2019). The indicator was that the qualitative data (created as a result of direct quotation as qualifying confirmation) coincide with each other. In addition, the activities used in the study were prepared within the framework of a plan accompanied by an expert in the field in both groups and this plan was implemented from start to finish.

The Participants

Since the study is a mixed-methods research, there is a need to collect qualitative and quantitative data from the participants. Therefore, we preferred purposive sampling in the study. According to Patton (2002), it provides ease of access to critical information for research purposes in addition to providing a wide range of information in terms of knowledge. Critical sampling was used in this study. In accordance with this method, students' participation in the physics laboratory course was chosen as the main criterion. Participants are undergraduate students of science education at one of the state universities in Turkey in 2017-2018 academic year. In this program, there are two cohorts (A and B) in which the students were assigned randomly into Class A experimental group with 42 (36 female, six male) students and Class B control group with 42 (40 female, two male) students. 84 students aged between 18-20 participated in the study. The technology itself can have a positive impact on the student's motivation to participate in the activity, but it can also be a limiting factor depending on competences and individual differences. (Nicol, Owens, Le Coze, MacIntyre, & Eastwood, 2018). Therefore, providing all students with technology may not produce the same results. For this reason, the technology literacy of the students in the experimental group was investigated. Table 1 shows that most of the students participating in the study were smartphone, Internet and Facebook users (5 years and more).

Table 1
Experiences by Participants Using Technology

Characteristic	f	%
Experience by students using smartphones		
1 year to 2 years	4	9.52
2 years to 4 years	16	38.10
5 years or more	22	52.38
Experience by students using the Internet		
1 year to 2 years	3	7.14
2 years to 4 years	12	28.57
5 years or more	27	64.29
Experience by students using Facebook		
1 year to 2 years	7	16.66
2 years to 4 years	16	38.10
5 years or more	19	45.24

Experimental Process

The study was conducted in accordance with the content of the general physics laboratory-II course. This course is a practical course with two lessons per week (45 min. + 45 min. = 90 min.). This course is guided by an instructor in the physics laboratory. The instructor is responsible for both groups and provides guidance and assistance to students in case of need. In both cohorts, a group of seven people is formed, and the applications are carried out by those students under the guidance of the instructor. Thus, the experiments were conducted in a laboratory through a collaborative method. The study was conducted in a six-week period during the 2017-2018 academic year. Information about the applications was given in the first week. In the remaining weeks, electromagnetism (experiments involving electric motor and electric bell operation principles), Ohm's law (experiment of measuring the resistance of a conductor), Wheatstone bridge (Determination of current point zero), Kirchhoff's law (Calculation of current and voltage separately) and Transformer (using primary and secondary windings) tests were done. In addition, these experiments are available in the general physics laboratory-2 course.

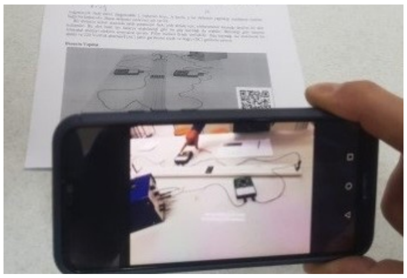
Control Group

A conventional method was used in the control group and the experiments were performed in a laboratory. In the first week of the application, experimental equipment was introduced to the students in the control group. In the remaining five weeks, the experimental manual including the experiments to be performed was given to the students. The experiment manual contains the following sections which students will follow in order: the name and purpose of the experiment, materials used in the experiment, brief theoretical information aiming to give information about the experiment, gradual information explaining how to conduct the experiment, findings for recording the obtained data, conclusion and discussion sections for generalizing the information and the evaluation section given for measuring performance (Open ended questions). The students in the control group perform their weekly experiments in the laboratory (in-class) through the guidance of an instructor. During the course of 90 minutes (45 + 45), they perform all the stages of an experiment in a laboratory in groups of seven participants.

Experimental Group

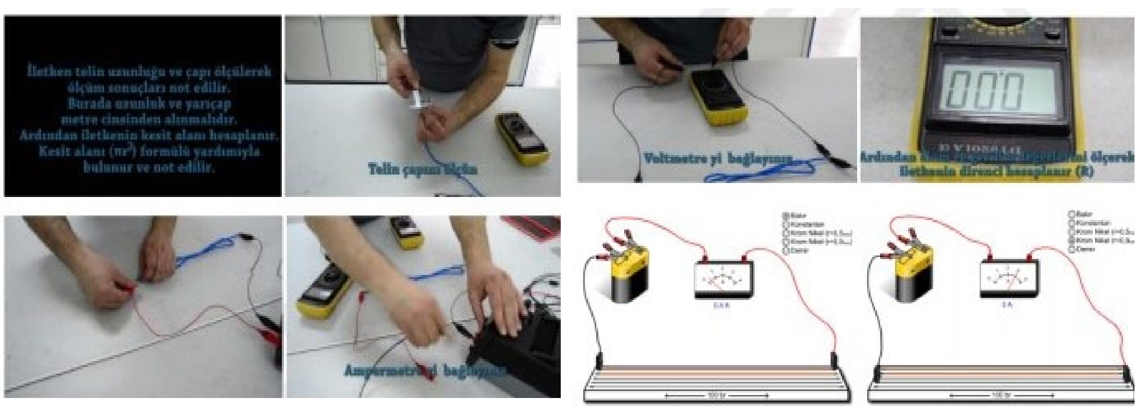
The experiments were performed both in laboratory (in-class) and out-of-school environments (out-of-class) in the experimental group using the FC model. In the FC model, teaching strategies such as collaborative group work in in-class applications are used to support students' learning. Therefore, the greatest benefit that the students perceive is to gain the ability to interact and collaborate with their peers. (McLean & Attardi, 2018). In the FC model, students should be able to use interactive materials in out-of-class environments as well as to use time effectively without having any communication problems with the instructor. Therefore, QR code applications have been added by the researchers on the experiment manual for interaction. The applications developed by the researchers for each experiment are presented to the students in connection with the QR in the AR-based laboratory manual of the experimental group. Links to video, graphics and additional content were used for incremental components in the applications. Take Ohm's law as an example (see Figure 1). When supported by the digital learning environment, students can access information at any time while working in a collaborative and engaging learning environment. Therefore, it is thought to provide a more effective education to students than the traditional method.

Figure 1
Students Interacting with QR Located in AR-based Laboratory Manual



In addition, researches that test the FC model show that it can lead to differentiation or reduction in student satisfaction and motivation at a later stage. The basis of this perception lies in the possibility of reducing the frequency and willingness to participate in activities contrary to the students' expectations. Therefore, when the studies in the literature are examined, one of the factors that can reduce the effectiveness of the FC model is the resistance by students to participate in activities. (Herreid & Schiller, 2013). In order to minimize the resistance, AR-based simulations and video applications have been developed to increase the motivation by students to participate in activities with smartphones anytime and anywhere. As a result of technology integration in accordance with the application of the FC model, the AR-supported laboratory manual, which was developed by the researchers, was used in the experimental group. Access to AR applications was closed when students arrived at the laboratory (in-class). In this way, in-class activities were conducted under the same circumstances for both groups. In the applications, there were 8-10 minutes of video on the installation of the experiment, 8-10 minutes of video introducing the experimental equipment to be used by the students and explaining the objectives, and a 12-15-minute video supported by simulations and animations describing the theoretical content of the experiment (see Figure 2). One of the simulated video applications developed by the researchers and used during the research process was given in Figure 2 (Ohm's law). These videos, which made an aggregate of 28-35 minutes, were given to students in the experimental group as out-of-class activities during the study.

Figure 2
Video Interaction and Simulation Applications in the AR-based Laboratory Manual



In the experimental manual for the experimental group, certain sections such as the experiment, the conclusion and the question and answer were conducted as in-class activities under the guidance of the instructor. Therefore, a total of 28-35 minutes was added in the experimental group in the process of the experiment. This additional time provides students with more time in the conclusion and question-answer sections of the experiment and helps them better understand the unclear and complex points under the guidance of the instructor. The time distribution of a 90-minute course was given in Table 2.

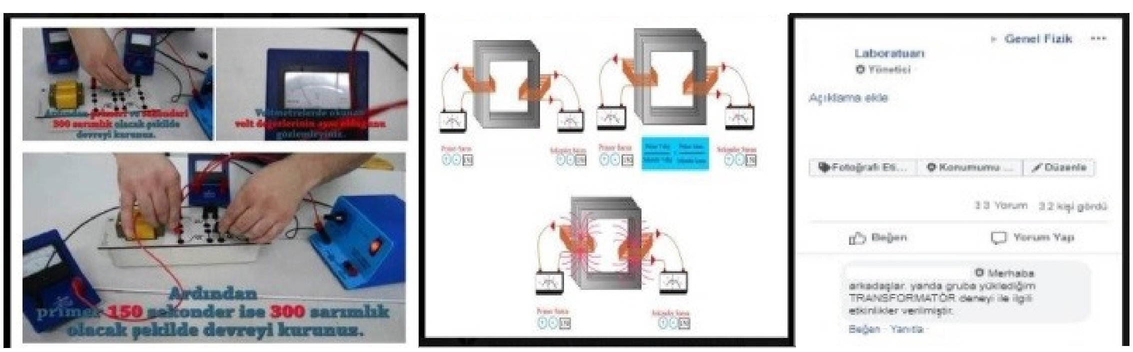
Table 2
Planning the Time Distribution of a 90-Minute Course

AR-based Laboratory Guide Sections	Flipped Classroom	Traditional Classroom
Aim Equipment Theoretical knowledge	Out-of-class activities (unlimited)	In-class activities (28-35 min.)
Experiment* Result Question-Answer	In-class activities (90 min.)	In-class activities (55-62 min.)
Total	90 min.	(28+62) = (35+55) = 90 min.

Note. This was carried out as both out-of-class activities (video watching for the experiment) and in-class activities (the experiment in the lab) for the experimental group.

As part of the out-of-class applications within the FC model, a Facebook group called the "General Physics Laboratory" was created in order to increase the interaction by the students in the experimental group with the instructor, to enable the students to use the time more effectively and to follow the work of the students (see Figure 3). Facebook is an online communication tool that allows users to create a custom profile for them to connect with or interact with people (Boyd & Ellison, 2007). Most undergraduate students at the university use Facebook on a daily basis (Kirschner & Karpinski, 2010). Students interact more on Facebook than online courses. In addition, some studies clearly reveal that there is the possibility that students can integrate Facebook into university courses (Bosch, 2009). Students believe that Facebook is a valuable resource as an academic tool, and that they are encouraged to develop academic connections and academic criticism, and to network and to improve the learning and learning experience (McCarthy, 2012).

Figure 3
Application Examples on Facebook



In addition, Rambe (2012) stated that Facebook has benefited students by encouraging the visibility of the common problems that students have with class-based concepts, and at the same time academics have easily recognized the difficulties faced by students. It is important to provide appropriate guidance and feedback for learning activities because feedback plays an important role in all learning activities. Therefore, the aim of the Facebook group is to motivate the students by creating a discussion environment with the questions and answers by the researchers. In this way, the researchers were guided and provided feedback to the students. The learning model and material used in the study were classified and summarized in Table 3.

Table 3
Learning Model and Material Used in the Application Process

Settings	Flipped Classroom Model (Experimental Group)		Traditional Model (Control Group)	
	Activities	Resources	Activities	Resources
Out-of-class (preparation before class)	- Preliminary study - Communication - Interaction	- AR-based laboratory manual - Facebook application	-	-
In-class (active learning)	- Conducting experiments in groups - Solve problem(s) - Peer discussions	- Laboratory manual - Guidance teacher	- Preliminary study - Conducting experiments by following laboratory manual sections	- Laboratory manual - Guidance teacher

Data Collection

Firstly, a protocol was given to the participants in the study to obtain data on participants' technology experiences; smartphone and Facebook usage times and the time spent on the Internet. In this study designed as a mixed-methods model, self-efficacy and attitude scales were used to collect quantitative data. A semi-structured interview form was used to gather qualitative data. In order to measure the students' laboratory self-efficacy, firstly, the "Criteria for the effective use of teachers' laboratory material" which was introduced by Linn and Gronlund (1995) into the literature and developed by Çepni et al. (2005) as "Items that measure teachers' attitudes and competences towards laboratory study" was adapted to the research problem. As a result of this study, the pre-test form of the data collection tool was formed by Böyük, Demir and Erol (2010). For each question in the questionnaire used in the research, the students were asked to rate as [(1) absolutely insufficient, (2) insufficient, (3) partially sufficient, (4) sufficient, (5) absolutely sufficient]. There were 18 questions in the questionnaire, so the self-efficacy scores of the students who participated in the study were a maximum of 90 points and a minimum of 18 points. Böyük et al. (2010) calculated the Cronbach alpha reliability coefficient of the scale as .89. The Cronbach alpha, which is the reliability coefficient of the questionnaire used in this study, was recalculated for both the pre-test and the final test and the new sample was taken into account. According to the data, the Cronbach's alpha reliability coefficient was found as .82 for the pre-test and .87 for the post-test.

For the validity of the scale, content validity study was conducted because the laboratory self-efficacy scale developed by Böyük et al. (2010) was developed for science teachers. Therefore, to adapt the scale to the pre-service teachers in higher education, eight instructors (science education experts) evaluated the items in the measuring instrument. The researchers asked the experts to evaluate the scale's suitability for pre-service teachers (by scoring 1 to 4 for each item). This evaluation was carried out with the content validity index on an item basis. According to the index, the degree of consistency between the raters (instructors) should be at least .80 (Szymanski & Linkowski, 1993). The compliance of the scale was found between .812 and 1.00 and 95.66% as a whole. The degree of consistency at this rate indicates an acceptable level. Inter-rater reliability was calculated by non-parametric Kendall's W- (Kendall's coefficient of concordance) analysis. The Kendall W test is used to determine the compatibility between more than two independent scoring (Legendre, 2005). As a result of this analysis, it was found that there was a statistically significant agreement between eight different experts for the 18 items. W (17) = .719 (X^2 = 97.812; p = .000). W <.70 and p> .05 can be considered high as the validity of the test (Legendre, 2005).

To determine the change in students' attitudes to physics laboratories before and after the application, the "Attitude Scale towards Physics Laboratory" which was developed by Nuhoğlu and Yalçın (2004) and tested for its reliability and validity in many studies in the related literature was used. The scale is rated with five options and four equally spaced Likert types ranging from "Strongly Agree" to "Strongly Disagree". The scale consists of 36 items related to university students' attitudes and perceptions towards physics laboratories. Nuhoğlu and Yalçın (2004) calculated the Cronbach's alpha reliability coefficient of the scale as .89. Likewise, the construct validity of the scale was provided by factor analysis by Nuhoğlu and Yalçın (2004). At the end of the rotation with Varimax Factor Analysis for scale items on attitude, it was decided that the scale was one-dimensional. Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett Sphericity test were used to determine the suitability of the data for factor analysis and sample adequacy. In this respect, the compliance of the data with the sample group was found to be KMO value at .854 and Barlet Test value was found as 3386.70. The sample is adequate if the value of KMO is greater than 0.5 (Field, 2000).

For the reliability of the scales, the stability of the measured values obtained from the repeated measurements under the same conditions is an important indicator (Carmines & Zeller, 1982). The Cronbach's alpha reliability coefficient of the questionnaire used in this study was re-calculated for both the pre-test and the final test, taking the new sample into consideration. According to the data, the Cronbach's alpha reliability coefficient was found as .78 for the pre-test and .81 for the post-test.

In order to reveal the opinions and suggestions of the participants in the experimental group on the use of the FC model in physics laboratories, a semi-structured interview form was developed and interviews were conducted with the students. The purpose of these interviews is to have an in-depth understanding of the students' thoughts on any activity.

The semi-structured interview form was given to the experimental group (N = 42) in which the AR-based FC model was applied. In this form, a question like "What are the advantages and disadvantages of the FC model teaching practices for the learning process?" would qualitatively describe the views of the participants in the experimental group.

Data Analysis

First of all, in order to gather information about the experiences of participants using technology, the time spent using the smartphone and Facebook and the time spent on the internet were analysed through the descriptive statistics method (see Table 2). An analysis of the collected data, including reliability analysis, was performed with a computer-aided statistical program. Before that, whether the data from the experimental and control groups showed a normal distribution was examined. In this study, Shapiro-Wilk test was used because the number of students was below 50 (n = 42). When the Shapiro-Wilk test for the "Laboratory Self-efficacy Scale" (experimental group, p= .17; control group, p= .22) and "Attitude Scale Towards Physics Laboratory" (experimental group, p= .13; control group, p= .34) was applied to both experimental and control groups, there was no significant difference in the p <.05 level in the experimental group and control group. Therefore, it was concluded that the experimental group and control group data showed normal distribution in both scales. For this reason, the parametric statistical tests were used in the analysis. Thus, an independent sample t test was performed as pre-test and post-test to determine whether there is a significant difference between the laboratory self-efficacy and laboratory attitude scores of the experimental and control groups.

An analysis of the qualitative data obtained as a result of the interview with the students who participated in the study on a voluntary basis was made by the content analysis method. Content analysis consists of a variety of processes for determining the research questions to be answered, selecting the sample to be analysed, identifying the categories to be applied, determining the coding process and coding training, applying the coding process, determining the reliability and analysing the results of the coding process (Kaid & Wadsworth, 1989). In the data analysis process, firstly the research data were coded and themes were formed. After that, codes and themes were arranged and the findings were interpreted. Students were coded in the form of Pre-service Science Teacher (PST1-PST42) numbered from 1 to 42.

With the help of expert opinions and objective coding, most content analysis studies can provide validity standard (Potter & Levine-Donnerstein, 1999). Reliability of the measuring instrument was tested with a percentage of the agreement between two field education experts (Şencan, 2005). Students' opinions were evaluated separately by two field education experts. Then, Reliability = Agreement / (Contract + Disagreement) reliability formula by Miles and Huberman (1994) was used in the content analysis of the data collected by calculating the matching ratios in the research. As a result, the reliability of the coding was calculated as 82% and considered to be reliable. The analysis of the research is expected to be more than 70% of reliability (Miles & Huberman, 1994).

Findings

RQ #1: *Is there a significant difference in terms of the laboratory self-efficacy between the experimental and control group?*

In order to test whether there is a significant difference between the laboratory self-efficacy scores of the students in both groups, a t-test analysis was performed. Thus, the equivalence of both groups was investigated and no significant difference was able to be found. After the application, the groups were given the same measurement tool as the post-test and the difference between the post-test scores of the groups was significant. Pre-test and post-test t test analyses are given in Table 4.

Table 4

T-test Results for the Pre-test and Post-test Means of the Experiment and Control Groups for Laboratory Self-Efficacy

Variable	Group	N	Mean	SD	T	p	Cohen's d
Pre-test	Control group	42	3.56	.21	1.52	.13	
	Experimental group	42	3.67	.34			
Post-test	Control group	42	3.77	.24	2.5	.014*	.545
	Experimental group	42	4.02	.59			

*<.05

Although the statistical significance tests are usually used to determine the difference between the mean scores of groups in the literature, it is not possible to say the same thing for the effect size which helps to make a more accurate decision about the results obtained by eliminating the effects caused by the number of samples in those significance tests. The effect size measurements calculated according to the difference of the group means were calculated as Cohen's d (Cohen, 1988), (.545) in Table 4 and (.713) in Table 5 in this study. The Cohen's d value obtained as a result of the calculations is interpreted as follows: .20- small effect size; .50- medium; .80 large effect size (Cohen, 1988). Table 5 reveals a significant difference in the post-test in favour of the experimental group. In addition, it was concluded that the FC model had a positive effect on the laboratory self-efficacy of university students.

RQ #2. *Is there a significant difference in attitude scores between the experimental and control groups?*

The attitude questionnaire was applied to both groups before and after the application to determine whether the FC model had had an effect on students' attitudes towards physics labs. As a result, no significant difference was found between the attitude scores of the groups before the application. Thus, it can be said that the attitudes of both groups are equal before the application.

Table 5
T-test Results for the Pre-test and Post-test Means of the Experimental and Control Groups for Attitude Scores

Varible	Group	N	Mean	SD	T	p	Cohen's d
Pre-test	Control group	42	3.67	.28	.65	.51	
	Experimental group	42	3.73	.52			
Post-test	Control group	42	4.00	.50	3.27	.002*	.713
	Experimental group	42	4.28	.26			

*<.05

At the end of the application, there was an increase in the attitude scores of the students in the experimental group. Table 5 shows a significant difference between the groups. As a result, the FC model seems to have a positive effect on students' attitudes towards physics laboratories.

RQ #3. *What is the opinion of the experimental group for the FC model?*

The data obtained from the interviews with the students are presented in two categories. The first one is classified as the positive effects of FC model and the other one as negative effects of the FC model. Therefore, when the data obtained from the interview were examined, the students in the experimental group thought that the FC model applied in the physics laboratory provided some positive and negative characteristics. Table 6 shows the coding and frequency distributions of the student expressions obtained from interview data according to the content analysis.

Considering the data in Table 6, some of the most remarkable statements of pre-service teachers are given below.

PST*23: "At first, while experimenting in laboratores, it was the most time-consuming process because it was a very difficult process to understand. But in the videos I watched in the FC practice, I completed the experiments in a very short time because I knew the tools used in the experiments and their purposes. So, I had more time to evaluate the other parts of the experiment."*

PST*12: "Generally, the question-answer section never really achieved its goal at the end of the experiments. These sections were either ignored or given to us as homework, but with this application more time remained in the laboratory to the question and answer section."*

PST*8: "Before these applications, I was attending classes using both internet and textbooks. But these methods were both boring and very useless. With this application, I wanted to participate more and more in activities without getting bored through the Facebook group and watch the videos prepared remarkably. I was also able to access the videos about the experiments at any time, so I had the opportunity to study anywhere and anytime."*

PST*2: "I accessed the videos prepared for us only at home, so watching all the videos took much time."*

Table 6
The Students' Comments (Advantages and Disadvantages of the FC Model)

Categories	Comments	Responses
Advantages of FC model	Experiments were completed in a shorter time	37
	More time was allowed for questions and answers at the end of the experiment	32
	I set the studying time myself	28
	Knowing the tools used in the experiment saved time	25
	Preliminary monitoring of the experiment facilitated it	19
	Videos were easily accessed	17
	As a member of the Facebook group I had the opportunity to study more efficiently	14
	I was afraid to ask questions, so I had the opportunity to watch videos again and again until I could understand them	13
Disadvantages of FC model	None	22
	Watching videos took a lot of time at home	10

Conclusion and Discussion

This study took place in a physics laboratory was conducted to measure the effects of the FC model on the laboratory self-efficacy and attitudes of university students and to find out about their opinions.

The experimental results regarding the first aim show that the FC model positively affects the laboratory self-efficacy of students. In accordance with the literature (Berrett, 2012; Deslauriers et al., 2011; Haak, HilleRisLambers, Pitre, & Freeman, 2011; Lin, 2019), this study shows that a class in which the FC model is applied can produce better learning outcomes as it increases the opportunity for more practice. Therefore, the experimental process helps students to acquire better laboratory self-efficacy (Aşıksoy & Özdamlı, 2016; Schultz et al., 2014).

Further to that, the data obtained regarding the second aim of the study show that the FC model positively affects students' attitudes towards the laboratory, which is also confirm by the related literature (See also Chao, Chen, & Chuang, 2015; Wanner & Palmer, 2015; Olakanmi, 2017). Studies on the FC practices reveal that students are more ready to participate in face-to-face, interactive and high-level activities such as problem-solving and discussion (Gaughan, 2014). However, the benefits claimed for the FC model are mainly reflected in the self-acquired perceptions and attitudes in the process rather than directly reflected in the learning outcomes (O'Flaherty & Phillips, 2015).

In contrast, Presti (2016) found that the FC model had no effect on students' attitudes. There are many studies that have negative findings as to the FC model providing a significant improvement in student performance (Blair, Maharaj, & Primus, 2016; Davies et al., 2013; Findlay-Thompson & Mombourquette, 2014; Karabulut-Ilgu, Jaramillo Cherrez, & Hassall, 2018). This situation was directly related to the change in the learning process and learning environment. Students may find the FC model uncomfortable at first sight, and some are not satisfied with the change in the traditional approach despite learning outcomes (Strayer, 2012). In addition, students may feel under pressure or anxious to complete their inverted learning activities (Marcum & Perry, 2015). Therefore, such negative consequences are possible. On the other hand, it is a different finding that initially students do not have a positive or negative view of preparing for the course in which the FC model is followed by watching online videos, but as the time progresses, students' attitudes towards the FC model have improved positively and they have adopted access to online content (Smallhorn, 2017). Moreover, although students' participation in the FC model, a very new learning environment, has been shown to lead to a satisfactory decrease in student attitudes (Gutwill-Wise, 2001), this practice has not adversely affected student satisfaction. The lack of negative attitudes in the study can be evaluated as a result of the AR and Facebook social media interaction as a measure in the learning process.

Also, student views regarding the third aim of the study revealed important results about the FC model. The experiments were completed in a shorter time, more time remained for the question & answer section at the end of the experiment, the time of the study was determined by the students, pre-understanding of the equipment used in the experiment saved time, pre-monitoring the experiment made it easier, videos were easily accessed, and as a member of the Facebook group, the students studied more efficiently in the group that the FC model was applied. The literature review suggests that the benefits of the FC model include improved participation, more student satisfaction, more class participation time, more flexibility, immediate support and feedback, higher-level thinking, and individualized learning experiences (Sams & Bergmann, 2013; Karamustafaoğlu, & Kılıç, 2018). Therefore, the study findings are consistent with the literature. In parallel with the findings of the study, Luo, Yang, Xue and Zuo (2019) concluded that the pre-class preparations, the teaching content and the exercises of the students were clearly useful in the classes in which the FC model was applied. Similarly, the study of Bedi (2018) found that students showed positive attitudes to video conferencing and 24-hour availability and participated in online learning in their spare time by using digital learning tools such as tablets and smartphones.

Furthermore, university students could easily access the videos with their smart phones and had the opportunity to watch them again with the help of AR applications used. Similarly, in the study conducted by Butt (2014), the students commented that the video sections including difficult subjects are frequently re-monitored and that re-watching the videos about a concept creates a very positive perception in them. Similarly, Wanner and Palmer (2015) have reported students' views that reflect the tremendous flexibility of the video's ability to identify their time with the FC concept in their study. In addition, Heijstra and Sigurðardóttir (2018) reports that in their study the flipped class has a distinct advantage over the traditional class. This advantage allows recordings to be displayed online, and students to view recordings several times, to pause for note-taking, to delete back if something is not clear for the first time, and to view recordings at their own pace. The opportunity to track recordings more than once can support or motivate students to self-control and learn the subject or concept. In addition, the use of AR and smart phones in the application has provided advantages to the students. Because it has been beneficial for students who hesitate to ask questions and have the opportunity to watch the videos until they understand them. In parallel with this finding in the research, students were found to be embarrassed for asking the concepts they did not understand in the traditional classroom environment. The findings of this study are similar to those of the literature. Students watch the videos again and again without considering the thoughts of other students in FC practices (Baepler, Walker, & Driessen, 2014; Schultz et al., 2014).

Students' views, which include the first two sub-problems of this study, show that the learning process open to both independent and interactive learning is recognized in connection with the digital and social media included in the FC model. In fact, Awidi and Paynter (2019) developed a five-point model that fostered learning in a digital learning environment including having access to resources and information, support elements and motivation, participation and collaboration, evaluation and feedback and finally the process of structuring knowledge (Awidi & Paynter, 2019). Therefore, the FC model is successful in self-efficacy and attitude with the elements considered in the study. In this study, the AR and Facebook applications intended to increase student motivation in the FC practice resulted in an increase in contrast to the decrease in motivation observed in the literature.

Considering students who expressed their views on the negative aspects of the FC model, the necessity of monitoring videos outside the classroom had a negative effect on students. They stated that watching videos took a lot of time at home. Therefore, they were negatively affected by the FC practices. If they do not follow the required courses before the lesson, they do not fully benefit from the FC practice which requires students to participate in the course (Butt, 2014). To eliminate this problem, it is necessary to create short videos and to get the opinions and suggestions of students. In addition, these applications should be frequently included in university classrooms so that students can become more familiar with them. Thus, there will be a significant increase in the positive attitudes and opinions of students towards the FC practices.

Limitations and Future Research

Considering the participants' opinions at the end of the research, it should be noted that the video duration should be short when preparing videos to be used in the FC model in order not to reduce the motivation of learners. Because they can get bored while watching videos that take a long time in out-of-classroom environments. When preparing videos for the FC model, an interview with a group of students will increase the efficiency of learning. Considering these situations, the disadvantages of FC model will be reduced. The study was limited to only two learning products for the physics laboratory. In the studies to be carried out for the FC model, the diversity of learning products can be increased and the results of applications in different disciplines can be examined. In addition, this study, which is limited by purposeful sampling, may provide a more generalization as a result of using different sampling models.

The authors received no financial support for the research, authorship, and/or publication of this manuscript.

Harun Çelik (haruncelik@kku.edu.tr) is an Associate Professor of Science Education at Kırıkkale University. His research interests include diversity in current and new approaches in science teaching, learning environments supported by information technology, laboratory approaches and science teaching, teacher training.

Hüseyin Miraç Pektaş (hmiracpektas@hotmail.com) is an Research Assistant of Science Education at Kırıkkale University. His research interests include diversity in learning environments supported by information technology, laboratory approaches and science teaching.

Orhan Karamustafaoğlu (orseka@yahoo.com) received his Bachelor of Science from the Department of Physics at Middle East Technical University in 1993. He received a PhD degree on Physics Education from Karadeniz Technical University in 2003. He is currently a Professor of Science Education at Amasya University. His research interests include new teaching methods for physics and science education.

References

Abeysekera, L., & Dawson, P. (2015). Motivation and Cognitive Load in the Flipped Classroom: Definition, Rationale and a Call for Research. *Higher Education Research & Development, 34*(1), 1–14. https://doi.org/10.1080/07294360.2014.934336.

Akbaş, O., & Pektaş, H. M. (2011, December). The effects of using an interactive whiteboard on the academic achievement of university students. *In Asia-Pacific Forum on Science Learning and Teaching, 12*(2), 1-19.

Akçayır, M., Akçayır, G, Pektaş, H.M. & Ocak, M.A. (2016). Augmented reality in science laboratories: The effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories, Computers in Human Behavior, 57, 334–342.

Akı, F. N., & Gürel, Z. (2016). Freshman year computer engineering students' experiences for flipped physics lab class: An action research. In: *AIP Conference Proceedings: Turkish Physical Society 32nd International Physics Congress (TPS32),* Bodrum, Turkey, 6-9 September 2016, pp. 070001. Turkey: AIP Publishing. https://doi.org/10.1063/1.4976422

Álvarez, B. (2012). Flipping the classroom: Homework in class, lessons at home. *Education Digest, 77*(8), 18–21. https://search.proquest.com/docview/1008666019?accountid=16369 Accessed 09 January 2019.

Aşıksoy, G., & Özdamlı, F. (2016). Flipped Classroom adapted to the ARCS Model of Motivation and applied to a Physics Course. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(6), 1589–1603. https://doi.org/10.12973/eurasia.2016.1251a

Awidi, I. T., & Paynter, M. (2019). The impact of a flipped classroom approach on student learning experience. *Computers & Education*, *128*, 269-283. https://doi.org/10.1016/j.compedu.2018.09.013

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education, 78*, 227-236. https://doi.org/10.1016/j.compedu.2014.06.006

Bandura, A. (1994). Self efficacy. In V. S. Ramachaudran (ed). *Encylopedia of human behavior*. New York, pp. 71-81.http://www.uky.edu/~eushe2/Bandura/BanEncy.html#top. Accessed 08 January 2019.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215. https://psycnet.apa.org/record/1977-25733-001. Accessed 08 January 2019.

Bedi, J. (2018). *A review of students' perceptions, engagement, and academic achievement in the flipped classroom*. Master thesis, University of Victoria, Canada. https://dspace.library.uvic.ca//handle/1828/9233. Accessed 09 January 2019.

Bergmann, J., & Sams, A. (2012). *Flip Your Classroom. Reach Every Student in Every Class Every Day*. Washington, DC: International Society for Technology in Education.

Berrett, D. (2012). How 'flipping'the classroom can improve the traditional lecture. *The Chronicle of Higher Education*, *12*(19), 1-14. https://www.chronicle.com/article/How-Flipping-the-Classroom/130857. Accessed 08 January 2019.

Blair, E., Maharaj, C., & Primus, S. (2016). Performance and perception in the flipped classroom. *Education & Information Technologies, 21*, 1465–1482. https://doi.org/10.1007/s10639-015-9393-5

Bosch, T. E. (2009). Using online social networking for teaching and learning: Facebook use at the University of Cape Town. *Communicatio: South African Journal for Communication Theory and Research*, *35*(2), 185-200. https://doi.org/10.1080/02500160903250648

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Böyük, U., Demir, S., & Erol, M. (2010). Fen ve teknoloji dersi öğretmenlerinin laboratuvar çalışmalarına yönelik yeterlik görüşlerinin farklı değişkenlere göre incelenmesi [Analyzing the proficiency views of science and technology teachers on laboratory studies in terms of different varriables]. *Turkish Science-Research Foundation, 3*(4), 342-349. https://dergipark.org.tr/en/download/article-file/799639. Accessed 08 January 2019.

Brewer, R., & Movahedazarhouligh, S. (2018). Successful stories and conflicts: A literature review on the effectiveness of flipped learning in higher education. *Journal of Computer Assisted Learning, 34*(4), 409-416. https://doi.org/10.1111/jcal.12250.

Burden, K., & Kearney, M. (2016). Future scenarios for mobile science learning. *Research in Science Education, 46*(2), 287-308. https://doi.org/10.1007/s11165-016-9514-1.

Burke, A. S., & Fedorek, B. (2017). Does "flipping" promote engagement? A comparison of a traditional, online, and flipped class. *Active Learning in Higher Education, 18*(1), 11-24. https://doi.org/10.1177/1469787417693487

Butt, A. (2014). Student views on the use of a flipped classroom approach: Evidence from Australia. *Business Education & Accreditation*, *6(*1), 33-43. http://www.theibfr2.com/RePEc/ibf/beaccr/bea-v6n1-2014/BEA-V6N1-2014.pdf#page=35. Accessed 08 January 2019.

Carmines, E. G., & Zeller, R. A. (1982). *Reliability and validity assessment.* Beverly Hills: Sage Publications.

Çelik, H., & Pektaş, H. M. (2017). Graphic comprehension and interpretation skills of preservice teachers with different learning approaches in a technology-aided learning environment. *International Journal of Science and Mathematics Education, 15*(1), 1-17. https://doi.org/10.1007/s10763-015-9667-9

Chambers, S. K., & Andre, T. (1997). Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *34*(2), 107-123. https://doi.org/10.1002/(SICI)1098-2736(199702)34:2<107::AID-TEA2>3.0.CO;2-X

Chao, C. Y., Chen, Y. T., & Chuang, K. Y. (2015). Exploring students' learning attitude and achievement in flipped learning supported computer aided design curriculum: A study in high school engineering education. *Computer Applications in Engineering Education, 23*(4), 514-526. https://doi.org/10.1002/cae.21622

Chen, G. D., Chang, C. K., & Wang, C. Y. (2008). Ubiquitous Learning Website: Scaffold Learners by Mobile Devices with Informationaware Technologies. *Computers & Education, 50*(1), 77-90. https://doi.org/10.1016/j.compedu.2006.03.004

Chen, Y., Wang, Y., & Chen, N. S. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education, 79*, 16-27. https://doi.org/10.1016/j.compedu.2014.07.004

Chuang, H.H, Weng, C. Y., & Chen, C.H. (2016). Which students benefit most from a flipped classroom approach to language learning? *British Journal of Educational Technology, 49*(1), 56-68. https://doi.org/10.1111/bjet.12530

Churchill, D., & Hedberg, J. (2008). Learning object design considerations for small-screen handheld devices. *Computers & Education, 50* (3), 881–893. https://doi.org/10.1016/j.compedu.2006.09.004

Cohen, J. (1988). *Statistical power analysis fort he behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Creswell, J. W. 2003. *Qualitative, Quantitative and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.

Çepni, S., Kaya, A., & Küçük, M. (2005). Fizik öğretmenlerinin laboratuvarlara yönelik hizmet içi ihtiyaçlarının belirlenmesi [Determining the physics teachers' in-service needs for laboratories]. *J. Turkish Educ. Sci, 3*(2), 181-196. http://dergipark.gov.tr/tebd/issue/26124/275193. Accessed 06 January 2019.
Davies, R. S., Dean, D. L., & Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Education Tech Research Dev, 61*, 563–580. https://doi.org/10.1007/s11423-013-9305-6
Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science, 332*(6031), 862-864. https://doi.org/10.1126/science.1201783
Er, E., Kopcha, T. J., Orey, M., Dustman, W. (2015). Exploring College Students' Online Help-Seeking Behavior in a Flipped Classroom with a Web-Based Help-Seeking Tool. *Australasian Journal of Educational Technology, 31*(5), 537-555.
Fidalgo-Blanco, A., Martinez-Nuñez, M., Borrás-Gene, O., & Sanchez-Medina, J. J. (2017). Micro flip teaching- an innovative model to promote the active involvement of students. *Computers in Human Behavior, 72*, 713–723. https://doi.org/10.1016/j.chb.2016.07.060
Field, A. (2000). *Discovering Statistics Using SPSS for Windows*. London: Thousand Oaks.
Findlay-Thompson, S., & Mombourquette, P. (2014). Evaluation of a Flipped Classroom in an Undergraduate Business Course. *Business Education & Accreditation, 6*(1), 63-71. https://ssrn.com/abstract=2331035. Accessed 06 January 2019.
Gaughan, J. E. (2014). The flipped classroom in world history. *The History Teacher*, *47*(2), 221-244. https://www.jstor.org/stable/43264225?seq=1#metadata_info_tab_contents. Accessed 06 January 2019.
Gutwill-Wise J. P. (2001). The impact of active and context-based learning in introductory chemistry courses: an early evaluation of the modular approach. *J. Chem. Educ., 78*, 684–690. https://doi.org/10.1021/ed078p684
Gençer, B.G. (2015). *A case study towards the implementation of the flipped classroom model in the schools.* Master Thesis. Bahçeşehir University, İstanbul.
Gómez-Tejedor, J. A., Vidaurre, A., Tort-Ausina, I., Molina-Mateo, J., Serrano, M. A., Meseguer-Dueñas, J. M., ... & Riera, J. (2020). Effectiveness of flip teaching on engineering students' performance in the physics lab. *Computers & Education, 144*, 103708. https://doi.org/10.1016/j.compedu.2019.103708
Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory Biology *Science*, *332*(6034), 1213-1216. https://doi.org/10.1126/science.1204820
Hao, Y. (2016). Exploring undergraduates' perspectives and flipped learning readiness in their flipped classrooms. *Computers in Human Behavior, 59*, 82–92. https://doi.org/10.1016/j.chb.2016.01.032
Hao, Y., & Lee, K. S. (2016). Teaching in flipped classrooms: Exploring pre-service teachers' concerns. *Computers in Human Behavior, 57,* 250–260. https://doi.org/10.1016/j.chb.2015.12.022.
Heijstra, T. M., & Sigurðardóttir, M. S. (2018). The flipped classroom: Does viewing the recordings matter? *Active Learning in Higher Education, 19(3)*, 211-223. https://doi.org/10.1177/1469787417723217
Herreid, C.F., & Schiller, N.A. (2013). Case studies and the flipped classroom. *Journal of College Science Teaching, 42*(5), 62-66. ttps://www.jstor.org/stable/43631584. Accessed 06 January 2019.
Hwang, G. J., Lai, C. L., & Wang, S. Y. (2015). Seamless flipped learning: a mobile technology enhanced flipped classroom with effective learning strategies. *Journal of Computers in Education, 2*(4), 449–473. https://doi.org/10.1007/s40692-015-0043-0.

Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE—Life Sciences Education, 14*, 1-12. https://doi.org/10.1187/10.1187/cbe.14-08-0129

Jones, V., & Jo, J. H. (2004). Ubiquitous learning environment: An adaptive teaching system using ubiquitous technology. In: *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (eds R Atkinson, C McBeath, D Jonas-Dwyer and R Phillips), Perth, Western Australia, 5-8 December 2004, pp. 468-474, Perth: ASCILITE. https://www.ascilite.org/conferences/perth04/procs/jones.html. Accessed 06 January 2019.

Kaid, L. L., & Wadsworth, A. J. (1989). Content analysis. In: Emmert P and Barker LL (eds) *Measurement of Communication Behavior*. New York: Longman, pp. 197–217.

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Hassall, L. (2018). Flipping to engage students: Instructor perspectives on flipping large enrollment courses. *Australasian Journal of Educational Technology, 34*(4), 123-137. https://doi.org/10.14742/ajet.4036

Karamustafaoğlu, O., & Kılıç, M.F. (2018). *A Content Analysis of The Researches In The Framework of Flipped Classroom Model*, 2nd International Symposium on Silk Road Academic Studies (ISSAS), The Book of Full Text p. 342-359, Nevşehir-Turkey.

King, B.M., & Newmann, F.M. (2001). Building school capacity through professional development: conceptual and empirical considerations. *International Journal of Educational Management, 15* (2), 86-94. https://doi.org/10.1108/09513540110383818

Kirschner, P. A., & Karpinski, A. C. (2010). Facebook® and academic performance. *Computers in Human Behavior*, *26*(6), 1237-1245. https://doi.org/10.1016/j.chb.2010.03.024

Legendre, P. (2005). Species associations: The kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics, 10* (2), 226–245. https://doi.org/10.1198/108571105X46642

Lin, Y. T. (2019). Impacts of a flipped classroom with a smart learning diagnosis system on students' learning performance, perception, and problem solving ability in a software engineering course. *Computers in Human Behavior*, 95, 187- 196. https://doi.org/10.1016/j.chb.2018.11.036.

Linn, R. L., & Gronlund, N, E. (1995). *Measurement and Assessment in Teaching*, (7th Edition), Prentice-Hall International (UK) Limited, London.

Luo, H., Yang, T., Xue, J., & Zuo, M. (2018). Impact of student agency on learning performance and learning experience in a flipped classroom. *British Journal of Educational Technology, 50*(2), 819-831. https://doi.org/10.1111/bjet.12604

Marcum, T. M., & Perry, S. J. (2015). Flips and flops: A new approach to a traditonal law course. *Journal of Legal Studies Education*, *32*(2), 255-286. https://doi.org/10.1111/jlse.12028

McCarthy, J. (2012). International design collaboration and mentoring for tertiary students through Facebook. *Australasian Journal of Educational Technology*, *28*(5), 755-775. https://doi.org/10.14742/ajet.1383

McLean, S., & Attardi, S. M. (2018). Sage or guide? Student perceptions of the role of the instructor in a flipped classroom. *Active Learning in Higher Education.* 1-13. https://doi.org/10.1177/1469787418793725

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks. CA: Sage Publications.

Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *Journal of Nursing Education, 52*(10), 597–599. https://doi.org/10.3928/01484834-20130919-03.

Moffett, J. (2015). Twelve tips for "flipping" the classroom. *Medical Teacher, 37*(4), 331–336. https://doi.org/10.3109/0142159X.2014.943710

Nicol, A. A., Owens, S. M., Le Coze, S. S., MacIntyre, A., & Eastwood, C. (2018). Comparison of high-technology active learning and low-technology active learning classrooms. *Active Learning in Higher Education, 19*(3), 253-265. https://doi.org/10.1177/1469787417731176

Nuhoğlu, H., & Yalçın, N. (2004). Fizik laboratuvarına yönelik bir tutum ölçeğinin geliştirilmesi ve öğretmen adaylarının fizik laboratuvarına yönelik tutumlarının değerlendirilmesi [The development of attitude scale for physics laboratory and the assessment of pre-service teachers' attitudes towards physics laboratory]. *Gazi University Kırşehir Faculty of Education Journal, 5*(2), 317-327. http://dergipark.gov.tr/download/article-file/15758. Accessed 06 January 2019.

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The İnternet and Higher Education*, *25*, 85-95. https://doi.org/10.1016/j.iheduc.2015.02.002

Olakanmi, E. E. (2017). The effects of a flipped classroom model of instruction on students' performance and attitudes towards chemistry. *Journal of Science Education and Technology*, *26*(1), 127-137. https://doi.org/10.1007/s10956-016-9657-x

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049-1079. https://doi.org/10.1080/0950069032000032199.

Owens, D. C., Sadler, T. D., Barlow, A. T., & Smith-Walters, C. (2017). Student motivation from and resistance to active learning rooted in essential science practices. *Research in Science Education*, 1-25. https://doi.org/10.1007/s11165-017-9688-1

Özmen, H. & Karamustafaoğlu, O. (2019). *Eğitimde Araştırma Yöntemleri* [Research Methods in Education], Ankara: Pegem Publishing.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*,543–578. https://doi.org/10.3102/00346543066004543

Palic, G., & Pirasa, N. (2012). A study of pre-service teachers' tendency for imprudent behaviour and physics laboratory attitudes. *Procedia-Social and Behavioral Sciences*, *47*, 823-828. https://core.ac.uk/download/pdf/82161005.pdf. Accessed 06 January 2019.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*, 3rd ed, Thousand Oaks. CA: Sage Publications.

Potter, W., J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*, 27, 258–284. https://doi.org/10.1080/00909889909655

Presti, C. R. (2016). The flipped learning approach in nursing education: A literature review. *Journal of Nursing Education, 55*(5), 252-257. https://doi.org/10.3928/01484834-20160414-03

Rambe, P. (2012). Critical discourse analysis of collaborative engagement in Facebook postings. *Australasian Journal of Educational Technology*, *28*(2), 295-314. https://doi.org/10.14742/ajet.875

Sams, A., & Bergmann, J. (2013). Flip your students' learning. *Educational Leadership, 70*(6), 16–20. http://www.ascd.org/publications/educational-leadership/mar13/vol70/num06/Flip-Your-Students'-Learning.aspx. Accessed 06 January 2019.

Samuel, L.M. (2019). Flipped pedagogy and student evaluations of teaching. *Active Learning in Higher Education*. https://doi.org/10.1177/1469787419855188

Schultz, D., Duffield, S., Rasmussen, S. C., & Wageman, J. (2014). Effects of the flipped classroom model on student performance for advanced placement high school chemistry students. *Journal of Chemical Education, 91*(9), 1334-1339. https://doi.org/10.1021/ed400868x

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik* [Reliability and validity in social and behavioral measurements], Ankara: Seçkin Publishing.

Sencar, S., & Eryılmaz, A. (2004). Factors mediating the effect of gender on ninth-grade Turkish students' misconceptions concerning electric circuits. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *41*(6), 603-616. https://doi.org/10.1002/tea.20016

Shih, W. L, & Tsai, C. Y. (2017). Students' perception of a flipped classroom approach to facilitating online project-based learning in marketing research courses. *Australasian Journal of Educational Technology, 33*(5), 32-49. https://doi.org/10.14742/ajet.2884

Shipstone, D. (1998). Pupils' understanding of simple electrical circuits. *Physics Education, 23*, 92–96. https://doi.org/10.1088/0031-9120/23/2/004

Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology, 66*(3), 531-568. https://doi.org/10.1111/peps.12035

Smallhorn, M. (2017). The flipped classroom: A learning model to increase student engagement not academic achievement. *Student Success*, *8*(2), 43-53. https://doi.org/10.5204/ssj.v8i2.381

Sohrabi, B., & Iraj, H. (2016). Implementing flipped classroom using digital media: A comparison of two demographically different groups perceptions. *Computers in Human Behavior, 60*, 514-524.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research, 15*(2), 171–193. https://doi.org/10.1007/s10984-012-9108-4

Szymanski, E. M., & Linkowski, D. C. (1993). Human resource development: An examination of perceived training needs of certified rehabilitation counselors. *Rehabilitation Counseling Bulletin, 37* (2), 163-176. https://doi.org/10.1891/0047-2220.24.4.58

Wanner, T., & Palmer, E. (2015). Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education, 88*, 354-369. https://doi.org/10.1016/j.compedu.2015.07.008

Wei, X., Weng, D., Liu, Y., & Wang, Y. (2015). Teaching based on augmented reality for a technical creative design course. *Computers & Education*, *81*, 221-234. https://doi.org/10.1016/j.compedu.2014.10.017

Wilson, S. G. (2013). The flipped class a method to address the challenges of an undergraduate statistics course. *Teaching of Psychology, 40*(3), 193-199. https://doi.org/10.1177/0098628313487461

Zepke, N. (2013) Threshold concepts and student engagement: Revisiting pedagogical content knowledge. *Active Learning in Higher Education, 14*(2), 97–107. https://doi.org/10.1177/1469787413481127

ELECTRONIC JOURNAL FOR RESEARCH
IN SCIENCE & MATHEMATICS EDUCATION
VOL. 25, NO. 2, 68-93

The Structural and Contextual Quality of Preservice Elementary Teachers' Argumentative Discussions

Evren Cappellaro
Akdeniz University

Demet Seban
Alanya Alaaddin Keykubat University

ABSTRACT

This study investigates the preservice elementary teachers' ability to use the components of an argument structure on a given science topic as well as the quality of the developed argument during their sophomore and senior year. In doing so, the study also aimed to discuss the effect of teacher education programs on the development of argumentation skills of students over two years. A qualitative research approach was applied. The data were collected from the same participants at the end of their second and fourth years through having them discuss different generic frameworks. The argumentative discourses were analyzed in four stages. The scientific argumentations' overall quality was at low-level. The participants' claims were generally established without using data and warrants. They failed to transfer their scientific knowledge into discussion using qualifiers and rebuttals. The quality of the data, claim, and warrants were insufficient and included misconceptions. The teacher education program increased the students' scientific knowledge to some extent. Courses with more argumentation-based discussions could help students to increase the quality of their argument and their capabilities to use components of an argument structure.

Keywords: preservice elementary teachers, argumentation, science education

Introduction

Today, the obtaining of first-hand information has become significantly easier. This development has placed emphasis on acquisition of scientific thinking skills rather than teaching of scientific knowledge and concepts in science education. Now, students are expected to use scientific knowledge, to understand its impact on their lives, to achieve conceptual learning through conducting studies and experiments, and subsequently to discuss and share what they have learned (Osborne et al., 2001). Increasing emphasis is placed on the teaching of argumentation-based science among the many other theories and approaches that have been developed in order to gain these skills. In recent years, the attention paid to the research in this field has been growing continuously (Erduran et al., 2015; Lee et al., 2009; Lin et al., 2014).

All forms of discussion enable students to acquire skills for information collection, selection, and inquiry and ensure their engagement in the social structuring of knowledge (Kuhn, 2005). However, argumentation differs in the process of transposition and hypothesis on learning. The importance of this approach comes from its practical aspects reflecting the rational thinking processes of scientists. That is why science learning requires argumentation skills as well as scientific knowledge. Argumentation-based scientific education facilitates meaningful learning for students. It helps students use cognitive/metacognitive strategies and processes, develops their communication skills, supports their critical thinking skills, promotes scientific literacy, and makes it easier for them to understand scientific culture and practice (Jimenez-Aleixandre & Erduran, 2008).

Teaching argumentation in science education also conveys an adequate image of science through constructing and analyzing arguments related to social applications and implications of science (Driver et al., 2000). This way, students can also realize that science and scientific knowledge is questioned, discussed, and subject to change. Argumentation does not only improve scientific thinking and inquiry skills but also the way students approach to the nature of science (Duschl & Osborne, 2002; Tumay & Koseoglu, 2011). Introducing students to this approach at an early age also contributes to their understanding of the nature of science and creates a positive impact on their approach to scientific concepts and sciences in future study (Ozdem et al., 2013).

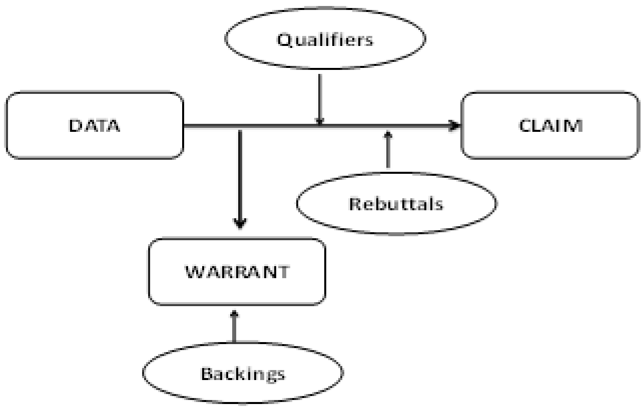
The legitimate answer to assess quality of arguments relies on the definition of argument and argumentation theories. Argumentation involves diverse meanings and many scientists have contributed to defining the concept of argument in the related literature in various ways (Jimenez-Aleixandre & Erduran, 2008; Kuhn, 1992; Plantin, 2005). Jimenez-Aleixandre and Erduran (2008) emphasized that the structuring of knowledge in the sciences is linked to evidence-based justification and that the claims should be presented within a logical context through its correlation to the data and evidence obtained from different sources. Therefore, they defined argument as constructing a connection between data and claims by means of rebuttal or evaluation in light of theoretical or empirical data in regard to a scientific topic/knowledge.

Toulmin made significant contributions to the development of the definition of scientific argumentation as a concept. He explains the structure of argument in a model consisting of four main elements (claim, data, warrant, backing) and two auxiliary elements (qualifiers, rebuttal) (Toulmin, 1958). According to this model, the knowledge and facts to support the claim is defined as *data*; the conclusion that is drawn is defined as *claim*; causes and principles to validate the relation between the data and claim is defined as *warrant*; and assumptions agreed to validate certain justifications is defined as *backing*. For more complex arguments, Toulmin then defined two more elements, namely qualifiers and rebuttals. *Qualifiers* are statements that limit the conditions under which the claim is true. *Rebuttals* are counterarguments or statements indicating circumstances when the claim is not true.

In examining the forms of argument, Walton (1996) defines argumentation schemes as abstract structures and frames them as dialogical. Many different argumentation schemes were distinguished based on their forms and content. Every argumentative scheme provides a set of questions for different types of reasoning such as inductive, deductive, defaultive, or abductive. They represent the structures of common types of arguments used in different contexts like everyday, legal, and scientific discourse. To analyze an argument, the argumentation scheme needs to be identified first. If the scheme is identified, the explicit and implicit features of the premises or the relation between premises and conclusion can be identified (Walton et al., 2008). Anderson et al. (2001) also point to the dialogical nature of the process of argumentation and emphasize the assumption of reasoning and representing contrasting perspectives. They argue that extended arguments can be broken down into recurrent patterns which they call "argument stratagems" (p. 2). A complete argument stratagem should include five different categories of information: the purpose, the conditions, the forms, the consequences, and the objections. Each of these stratagems depends on an argument schema that helps to organize information, retrieve relevant information from memory, facilitate argument invention, provide the basis for anticipating objections, and find flaws in the arguments.

According to Sampson and Clark (2008), a great number of diverse analytical and conceptual frameworks are used to analyze argumentation processes carried out in research done in relation to argumentation-based education in science education. According to the researchers, the models developed by Toulmin (1958) and Schwarz et al. (2003) were two different approaches employed in research investigating the arguments constructed by students. The common aspect of these two approaches is that they can easily be generalized and adapted to diverse research fields. Toulmin's (1958) argument pattern is one of the most common argument models employed (Figure 1). In this model, what is investigated is usually the argument elements and not the content of an argument.

Figure 1
Toulmin Model of Argument



According to Toulmin's model, arguments that are deemed as strong can sometimes fall short of scientific content. As the presence of argument components cannot be an indicator of a qualified and well-established argument, Toulmin's model is used together with other supporting approaches when conducting research (Sampson & Clark, 2008). The framework of Schwarz et al. (2003), on the other hand, is for examination of arguments that have been developed in written discourse. This framework is used mainly for analyzing the argument in terms of content.

The discussions on how argumentation is situated in science and how it intervenes in science education legitimized argumentation-based science education. Different integration of argumentation as a part of curriculum and teacher training for the implementation of argumentation in science education became two aspects to consider for researchers who work in this area. Most national curricula expect students to acquire general argumentative skills, and science educational policies across the world highlight argumentation as a process of scientific knowledge construction (Erduran & Jimenex-Aleixandre, 2008). In Turkey, there were no standards directly related to argumentation in elementary school programs until 2015. "An interdisciplinary perspective based on an inquiry-based learning approach" (MEB, 2015) was adopted and argumentation-based science teaching was explicitly highlighted in the science curriculum beginning in elementary grades. Adopting a new revised curriculum requires time, and the success of implementation depends on the teachers' ability to implement argumentation practices. Besides providing in-service teachers with the necessary training, teacher education programs should consider new developments in the curriculum and revise their own curricular activities according to these changes. In addition to subject matter knowledge, preservice teachers are also educated on different instructional approaches and teaching techniques. The instruction that involves higher order thinking should be specifically designed to actively engage students to learn the process of thinking as scientific argumentation skills do not naturally develop and argumentation is not a process that spontaneously transpires in the class (McNeill & Pimentel, 2010).

The majority of studies carried out in relation to the use of argumentation in science education were done on preservice science teachers, inservice teachers at the level of middle school and high school, or students studying at this level of education (Cavagnetto, 2010). The number of studies on primary school students and teachers, on the other hand, is scarce in comparison to these studies (Cavagnetto et al., 2010; Kim & Hand, 2015). Therefore, preservice teachers who teach science should be equipped in terms of argumentation-based teaching. The effective use of this approach by the elementary level preservice teachers who are to guide students in argumentation-based science teaching and encourage meaningful learning depends on the fact that they understand the nature of scientific knowledge and that they have the necessary knowledge and skills for implementing this approach. This is also a necessity because success in middle school depends on how well students were introduced to argumentation-based reasoning and discussion.

Given this theoretical and empirical background to the problem, this study concerns preservice elementary teachers' scientific argumentation development from the second to fourth year of their program. In doing so, the study also aimed to discuss the effect of teacher education programs on the development of argumentation skills. Therefore, the research questions are:

- What are the components of argument structures of scientific discussions developed by sophomore and senior preservice elementary teachers?

- What is the quality of arguments developed by sophomore and senior preservice elementary teachers?

- To what extent does the structural and contextual quality of preservice elementary teachers' scientific argument develop from sophomore to senior year?

Method

This study aims to examine preservice elementary teachers' ability to use argumentation structure and the quality of the developed argument content. It also has the objective to discuss the contribution made by the education received over two years to the structural and contextual quality of preservice elementary teachers' scientific argument. The study employs a qualitative research method as it provides a holistic picture of the phenomena to understand the research problem (Cresswell, 2007). Qualitative research aims to draw a comprehensive picture and to interpret the meanings inferred from the phenomenon (Denzin & Lincoln, 1998).

Participants

The study was administered to 33 preservice elementary teachers. The participants are the same students who studied during the academic year of 2013-2014 in their second year and the academic year of 2015-2016 in their fourth. Seventeen female and 16 male students were voluntarily involved in the study. Purposeful sampling techniques were used to select participants in which the researcher selected participants who were associated with the research problem being studied (Creswell, 2007). The inclusion criteria were their success in science lessons in the first and second years of the education program.

Data Collection

With a view to examine the preservice elementary teachers' argumentation-based scientific discussions and the change that occurred in the structure and the content of these arguments, they were given a separate appointment at the end of the final exams so as to collect the data on two different occasions. The first data collection was carried out after the completion of year two by the students. Year two was chosen because the preservice teachers completed General Biology, General Physics, Chemistry, Instructional Principles and Methods, Science and Technology Laboratory Practices I and II at the end of that academic year. The students are expected to be knowledgeable about the basic concepts of science at the end of this year. Two years later, the second data collection was conducted right before the participants graduated. The students completed Science and Technology Education I and II, Life and Social Studies Education, and School Experience and Teaching Practices I and II throughout the last two years of the program. The preservice teachers gained experience as to how to convey knowledge and skills with regard to the science content they had learned. The students did not get any specific intervention program, nor were they trained specifically about argumentation using Toulmin's argumentation model. This was not preferred because teacher education programs are expected to provide students with necessary knowledge that they will need to make arguments.

Five groups of preservice teachers were formed, and each group received one generic question. When the groups were established, partner selection choices of the participants were taken into account and a group consisted of at least six students.

Data Collection Tool

The questions distributed to the groups were selected from among the questions developed by Osborne et al. (2001) with the contributions of a group of teachers. The researchers designated various argument types and gave exemplary discussion questions for each type of argument to improve the scientific argumentation skills of children. The questions were translated from English to Turkish and used after being revised in line with the opinions of two expert science teachers (Table 1).

Data Analysis Method

The group discussions were audio-recorded and lasted about 20 minutes. The discussion transcripts were analyzed in four stages. During the analysis process, categories that were pre-determined were used for each stage.

In the first stage, all data were analyzed by using Toulmin's (1958) argument pattern. The dimensions constituting this model are claim, data supporting the claim, warrants indicating the relation between the data and the claim, backing strengthening the warrants, qualifiers, and finally the rebuttals showing the conditions or events under which the claim is invalid. This stage of the study was conducted to determine which dimensions of the argument were used by the preservice teachers. The analysis of the data was undertaken by two researchers. One of the researchers has Ph.D. in science education and the other in Curriculum and Instruction in elementary education. One researcher is an expert in teaching elementary and middle school science as the other works within literacy and is very familiar with teaching how to write argumentative essays. General evaluations were noted down after having listened to the interview records. Subsequently, the researchers identified the claims, warrants, backings, qualifiers, and finally assumptions used through scrutinizing the written texts and then checked whether the discourses were coded into similar categories. The discourses that were not coded similarly were discussed and evaluated under the related category upon agreement.

In the second stage, the content of the argument was evaluated in terms of strength and weakness. In this evaluation, the framework developed by Okada and Buckingham Shum (2008) was used (Table 2). In this stage, the data analysis was conducted without considering the qualities of argument elements, only taking into account their presence and absence. In the tables, "2Y" represents second year students, "4Y" represents fourth year students. The subsequent number represents the group and the letter represents the name assigned to the preservice teacher in that group.

Table 1

Materials for Argumentation

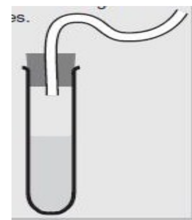
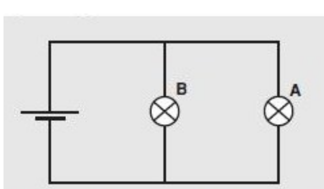
The Group	Generic Frameworks
1.	**Competing theories** **Theory 1**: Microbes are made from rotting material. This is helped by flashes of lightning, bright light and warmth. **Theory 2**: Microbes are carried in the air, probably on dust, and cannot be made out of dead matter. *Discuss whether the following pieces of evidence support Theory 1, Theory 2, both or neither.* **a.** Boiled soup in a sealed glass flask will keep forever. **b.** If cartons of milk are opened, they will not stay fresh. **c.** If the air was full of microbes, you wouldn't be able to see at all. **d.** If milk is heated and sealed, it will keep for several days. **e.** Boiling any material kills the vital ingredients needed to make microbes. **f.** Boiled soup, exposed to the air with a special S bend tube in it like the one shown here, does not go off. **g.** Food goes off more quickly in the summer when it is warm and humid. 
2.	**Constructing an argument** Heavier things do not always fall faster. *Look at the following statements of evidence. Discuss them with the others in your group and put them in a logical order to justify the statement above.* **a.** A penny and a brick reach the ground at the same time when dropped from the same height. **b.** Air resistance is a force which opposes motion. **c.** All things fall at the same rate if you ignore air resistance. **d.** A piece of paper falls much more slowly than a brick.
3.	**Understanding an argument** *Which of the following arguments provide good evidence that matter is made up of particles, and why?* **a.** Air in a syringe can be squeezed. **b.** All the crystals of any pure substance have the same shape. **c.** Water in a puddle disappears. **d.** Paper can be torn into very small pieces.
4.	**Experimental data** Everybody in the class measured the boiling point of water. They obtained the following results. 96 ˚C, 94 ˚C, 102 ˚C, 106 ˚C, 108 ˚C, 92 ˚C, 101 ˚C, 86 ˚C, 97 ˚C, 103 ˚C *In your group discuss:* **a.** Why might they disagree? **b.** How might they agree on a value?
5.	**Predicting, observing and explaining** Bulb A and Bulb B are two identical bulbs. *What will happen to the brightness of lamp B when lamp A is unscrewed?* *Discuss in your group and give reasons for what you think will happen.* 

Table 1
Categories of Argument Content

Argument Level	Statement	Example
Very weak	Only claim	*I think the pencil would fall faster.* (2Y-2C)
Weak	Claim and warrants (usually based on beliefs)	*As the density of air is close to that of the paper, the fall would slow down at this point.* (2Y-2D)
Moderate	Claim, (weak) warrants, data, or rebuttals	*Volume has effect. If the volume is bigger, then the frictional force applies more power.* (2Y-2A)
Strong	Claim, warrants, rebuttals, and/or data	*The volume of brick is bigger. Of course, if it is a full brick. As the volume of the brick is bigger, more frictional force applies.* (2Y-2C)

In the third stage, the claims of the arguments were evaluated as being right, wrong, or having misconceptions, whereas the warrants and data were evaluated as being sufficient, insufficient, and having misconceptions (Table 3).

Table 2
Categories of Argument Quality

Categories	Aspect	Example
Right claim	Scientifically right	*I think this can stay as it is forever because it has no contact with air. (4Y-1D)*
Wrong claim	Scientifically wrong	*I think the pencil would fall faster.* (4Y-2C)
Claim with misconception	Partially scientifically right, but it has misconceptions	*It is not in liquid phase at 103 degrees centigrade.* (2Y-4B)
Sufficient warrants	It supports the claim in every aspect.	*I begin first. Water, pure water boils at 100 degrees centigrade under atmospheric pressure. The characteristics of water, its purity, and its location and therefore its atmospheric pressure, they all change the boiling point. So, the differences in measurements may be due to these. The purity of water may be disrupted, changed or the atmospheric pressure where the boiling point is being measured may be increased or decreased. Thus, for the option A, I believe the different measurements may be the consequences of these changes.* (4Y-4E)
Insufficient warrants	It supports the claim in every aspect but is scientifically weak.	*As you know, the air in it can be compressed and it has particles inside, so that is why we can compress it.* (2Y-4B)
Warrants with misconception	It supports the claim partially but has misconceptions.	*The speed of putrefaction decreases related to working of enzymes.* (2Y-1A)
Insufficient data	Data supporting the claim only from one aspect.	*Ultimately, the matter itself is made of atoms, isn't it?...* (2Y-3B)
Data with misconception	Data having misconceptions.	… *These particles should be flexible or should shrink.* …(4Y-3F)

In the fourth stage, the data collected at the end of the second and fourth year were compared considering the aforementioned dimensions.

Results

In this phase, the arguments constructed by all the preservice teachers who participated in the research were first evaluated in terms of argument components and then their qualities. The groups were compared overall under these two elements.

Basic Components of Constructed Scientific Arguments

The arguments structured by the preservice teachers engaged in the research during the discussions were analyzed according to the aforementioned evaluation criteria. When the arguments are compared as frequency values, 127 arguments were developed in total at the end of year two, whereas this number increased to 139 at the end of year four. The highest number of arguments developed at the end of year two belonged to the second group discussing free fall. Meanwhile, the groups in year four developed almost equal numbers of arguments in general. The fifth group developed the lowest number of arguments when asked to construct arguments in relation to electric circuits at the end of years two and four (Table 4).

Table 4

Percentage Distribution of Argument Components in Group Discussions

Categories	Groups											
	1		2		3		4		5		Total	
	2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y
	%	%	%	%	%	%	%	%	%	%	%	%
Very weak	5.5	7.1	14.9	5.7	5.5	10	13.3	5	0.7	0	40	28
Weak	7.8	12.2	11	6.4	7.8	5	3.9	10	2.3	4.3	33	38
Moderate	3.9	1.4	7.8	10.7	6.2	8.6	1.5	4.3	3.9	4.3	24	30
Strong	0	0	2.3	0	0.7	0.7	0	2.1	0.7	0.7	3	4
Total	17.3	20,7	36	22.8	20.2	24.3	18.7	21.4	7.6	9.3	100	100

When Table 4 is examined, according to Toulmin's argumentation model, it was observed that the majority of the arguments developed at the end of year two involved arguments that were very weak and weak, and at the end of year four, weak and moderate. While very weak arguments were as high as 40% at the end of year two, this number decreased to the level of 28% at the end of year four. Moderate level arguments were equal to 25% of all the arguments developed at the end of year two, while this increased to 30% at the end of year four. The number of strong level arguments were similar at the end of years two and four.

When the groups were compared according to the argument categories, it was seen that the highest number of very weak arguments belonged to the second and fourth groups at the end of year two. While the rate of arguments of these groups under the very weak category decreased at the end of year two, this rate increased for the first and third groups. When, however, the weak arguments composed of claims and warrants developed in all the groups were compared, it was concluded that at the end of year two, the second group created more weak arguments in comparison with other groups, and at the end of year four, the first group created more weak arguments in comparison with other groups. As for the fifth group, they produced a smaller number of arguments in this category as compared to other groups both at the end of year two and year four.

The highest number of moderate level arguments was produced by the second group. When all the arguments at the moderate level were compared, the rate of these arguments produced by the first group decreased at the end of year four, whereas it increased in other groups. It was seen that the first and fourth groups did not form any strong arguments at the end of year two when the strong arguments developed by all the groups were examined, whereas the second group created a higher number of arguments in this category. At the end of year four, no arguments were detected in this category, either in the first or second group. While there were arguments present in this category in the discussion carried out by the second group at the end of year two, they were not able to produce any strong arguments at the end of year four. Although the fourth group did not make any strong arguments at the end of year two, they did produce strong arguments at the end of year four. As for the other groups, the rate of arguments in this category did not change.

Understanding the Nature of Scientific Argument: Competing Theories

In this category, it was observed that most of the participants discussing two conflicting theories in relation to microbes developed very weak and weak arguments at the end of years two and four (Table 5). The rate of weak arguments increased at the end of year four. The arguments with claims and rebuttals were present at the end of year two, while there were none at the end of year four. In comparison to year two, year four arguments comprised all of the following: claim, warrant and rebuttal. As for the moderate level arguments in this category, while the arguments in which claim and rebuttals stand out at the end of year two, no arguments in this category were observed to have been formed at the end of year four.

Table 5
Percentage and Frequency Distribution of Understanding the Nature of Argument Components

		2Y		4Y	
	Categories	f	%	f	%
Very weak	Claim	7	32	10	33
Weak	Claim+ warrant	10	46	17	57
Moderate	Claim+ data	1	5		
	Claim+ rebuttals	3	14		
	Claim+ data+ warrant	1	5	2	7
	Rebuttal				
Strong	Claim+ warrant+ rebuttal			1	3
	Claim+ data+ warrant+ rebuttal				
Total		22	100	30	100

The data used in the discussion is the data that were created based on the information which was embedded in the question itself. While only one participant used data in the arguments at the end of year two, two different participants made use of warrants and claims in addition to data at the end of year four. Furthermore, another participant developed an argument that contained claim, warrant, and rebuttal:

> *I think this can stay as it is forever as it has no contact with the air because microbes are transmitted through air. However, if air is present or leaks in the ambient, then it may not endure.* (4Y-1D)

Constructing an Argument: Free Fall

When participants constructed arguments on the claim that heavy objects do not always fall fast were analyzed, it was observed that while very weak arguments were formed at the end of year two, it was the moderate-level arguments with the highest number obtained at the end of year four (Table 6). The rate of very weak arguments decreased at the end of year four. When the category with moderate arguments is examined, what is striking is that the participants in this group used rebuttals or data together with the claim. While no arguments were detected having claim, data, and warrant at the end of year two, such arguments in this category had the highest number at the end of year four. Strong arguments were identified at the end of year two; however, the participants did not construct any strong arguments at the end of year four. There were participants in the group who could not form any arguments at all.

Table 6

Percentage and Frequency Distribution of Argument Components for Constructing an Argument

		2Y		4Y	
	Categories	f	%	f	%
Very weak	Claim	19	42	8	25
Weak	Claim+ warrant	14	31	9	28
Moderate	Claim+ data	2	4	2	6
	Claim+ rebuttals	3	7	4	13
	Claim+ data+ warrant			6	19
	Rebuttal	5	11	3	9
Strong	Claim+ warrant+ rebuttal	1	2		
	Claim+ data+ warrant+ rebuttal	1	2		
Total		45	100	32	100

Although during the discussion the participants mentioned in their arguments the concept of mass in relation to free fall in general, they included too many unrelated concepts such as density, volume, and weight. Moreover, not one among the participants used the most important concepts of free fall; that is, gravitational acceleration, cross-section, air resistance, or friction:

> *The one with more mass would fall faster.* (2Y-1D)

> *Let's not forget that its density is different.* (4Y-1A)

Understanding an Argument: Structure of Particulate Matter

When Table 7 is examined, it can be seen that the highest numbers of argument categories for the four claims about the particulate matter's structure were weak and moderate arguments at the end of year two and very weak and moderate arguments at the end of year four. While weak-level arguments were lesser in the group discussion at the end of year four, very weak arguments increased in number. Meanwhile the rate obtained for strong arguments showed no change. The claims were mostly presented with data in the moderate arguments at the end of year two. At the end of year four, in addition to arguments in which claim and data were used together, arguments with claim, data, and warrants were detected:

> *We know from the empirical evidence that syringe can be pressurized, in other words, as the distance of particles in gas phase is larger, they can be compressed.* (2Y-3B)

> *We are saying that this would apply for every matter, right? But, let's consider pebbles instead of sand. Pebbles can, too, accumulate and when it does and when we pour water in it, the water would flow through the pebbles. They would not hold anything.* (4Y-3E)

Table 7
Percentage and Frequency Distribution of Argument Components for Understanding an Argument

		2Y		4Y	
	Categories	f	%	f	%
Very weak	Claim	7	27	14	41
Weak	Claim+ warrant	10	39	7	20
Moderate	Claim+ data	4	15	5	15
	Claim+ rebuttals	2	8	3	9
	Claim+ data+ warrant	2	8	4	12
	Rebuttal				
Strong	Claim+ warrant+ rebuttal	1	4	1	3
	Claim+ data+ warrant+ rebuttal				
Total		26	100	34	100

Albeit similar to the arguments in the discussion, the only strong argument identified at the end of years two and four involved claim, warrant, and rebuttal. When examined scientifically, the warrant supported the claim from only one aspect. For instance, the most striking characteristics of the strong argument presented at the end of year two was that no scientific terminology was used except for the one provided in the question:

> *If it was not a particulate matter, if water was in bulk, it could not get through the soil or evaporate into the air as a whole; thus, this shows that it has a particulate structure.* (2Y-3E)

The group members failed to present counter-claims or did not address the validity of the claims as they did in other questions. The claims lacked scientific concepts and usually focused on options A and C. Although one participant was curious about whether the particles would differ for option B, s/he failed to turn it into a claim. And the question asked by this participant was not answered by other participants. This was also the case when other options were addressed during the discussion process. Instead of providing an answer to the claims presented, the participants sometimes changed the subject at hand, which created an environment in which the group conducted a disorganized conversation-like session. In this group's discussion, the scientific concepts were mentioned much less than the previous group.

Interpreting Experimental Data: Boiling Point of Water

The fourth group was asked to elaborate on the reasons for different results obtained from an experiment in which water's boiling point was measured. When this discussion was examined, it was seen that very weak and weak arguments were produced at the end of year two, and weak and moderate arguments were formed at the end of year four. While, the number of very weak arguments lessened at the end of year four, the number of moderate arguments increased. No strong arguments were detected at the end of year two, whereas at the end of year four, strong arguments were present (Table 8).

Table 8
Percentage and Frequency Distribution of Argument Components for Interpreting Experimental Data

	Categories	2Y f	2Y %	4Y f	4Y %
Very weak	Claim	17	71	7	23
Weak	Claim+ warrant	5	21	14	47
Moderate	Claim+ data			3	10
	Claim+ rebuttals	2	8	1	3
	Claim+ data+ warrant			2	7
	Rebuttal				
Strong	Claim+ warrant+ rebuttal			2	7
	Claim+ data+ warrant+ rebuttal			1	3
Total		24	100	30	100

The highest number of very weak arguments produced among groups was for this question; that is, discussion where empirical data were to be interpreted in relation to the boiling point of water. The participants made consecutive claims during the discussion:

The water could be calcareous or salty … could be a student error. (2Y-4A)

When the group discussion was analyzed, it was observed that the participants did not hold a discussion on the impact of the claims presented on the data in the question. When, however, the claims were examined, as opposed to the structures of the claims presented by other groups, it was seen that the participants in this group did not use simple present tense, rather they used modal verbs of probability:

It could be fresh water. (4Y-4B)

The group focused on a mixture of student error or measurement error rather than the concepts of height and pressure. As for option B in the question, at the end of year two, they recommended calculating the arithmetic average after summing up all related values. Only one participant (4D) objected to this and recommended repeating the experiment under a supervisor but the calculation of arithmetic average was approved by the other participants. However, when this claim was presented again at the end of year four, it was rejected by other participants:

When we discussed this the last time, I remember very well that we had said that we could calculate the arithmetic average of all the values. (4Y-4C)

Predicting and Explaining: Parallel Electric Circuit

The lowest number of arguments constructed among the groups was the discussion where the group addressed the results of loosening the connection of two parallel bulbs attached to an electric circuit (Table 9). Although very few arguments were formed, they were superior in strength when compared to those created by other groups. While the moderate arguments were almost equal to half of the total number of arguments, they usually composed claim, data, and warrant. The strong arguments constructed at the end of year two included elements of data, warrant, and rebuttal:

> *I believe no changes will be observed because they are both identical and the same current flows through both. If they were switched on, they would glow in the exact way. They would not be brighter than one another. Loosening A would not mean to increase the brightness of B.* (2Y-5D)

Table 9
Percentage and Frequency Distribution of Argument Components for Predicting and Explaining

		2Y		4Y	
		f	%	f	%
Very weak	Claim	1	10		
Weak	Claim+ warrant	3	30	6	46
Moderate	Claim+ data	2	20	2	15
	Claim+ rebuttals				
	Claim+ data+ warrant	3	30	4	31
	Rebuttal				
Strong	Claim+ warrant+ rebuttal			1	8
	Claim+ data+ warrant+ rebuttal	1	10		
Total		10	100	13	100

At the end of year four, data were included in the strong arguments:

> *I think it would not change because the current flowing through both parallel identical bulbs is the same. For instance, the bulbs at home are connected in parallel and when one is switched on, the other's brightness does not dim, right? But, if they were connected in series, then its brightness would dim.* (4Y-5C)

Among the group discussions considered to be the shortest was the one conducted for this question. The participants in this group carried out a discussion where no one undertook the effort to refute or support each other's ideas, although many ideas were brought forward, which was the case for the third group as well. Each member of the group participated in the discussion equally. Concepts of voltage and impedance used for the explanation of electric circuits were not mentioned in the discussion.

Basic Qualities of Constructed Scientific Arguments

Another objective of the research was to examine the qualities of argument elements developed by the participants at the end of years two and four (Table 10). Numbers of claim, warrants, and data were examined at the end of years two and four, respectively: 122 and 136 claims; 50 and 76 warrants; and 13 and 30 data. The highest number of claims was obtained in the second group's discussion on free fall with 40 and 36 claims at the end of years two and four, respectively. While the second group included the highest number of warrants (16 warrants) in their discussion at the end of year two, it was the first group that reached the highest number of warrants with 19 warrants at the end of year four. The arguments with the highest number of data were made by the participants in the third group. The first and fifth groups used the least of data by using data in only one argument.

When Table 10 is considered overall, it can be deducted that the preservice teachers at the end of year four made claims that were scientifically more accurate and their misconceptions lessened; that a sufficient number of warrants, although not many, was present and the misconceptions in warrants lessened; and that adequate level of data were included.

When their choice of claim was examined, it can be stated that while some of the claims were right, some of them were wrong and some of them had misconceptions. The second group who produced the highest number of claims presented more scientifically wrong claims with more misconceptions as opposed to other groups. However, overall, they produced correct claims at the end of years two and four. Only the fifth group at the end of year two presented claims of which 80 percent was wrong scientifically:

> *I believe that since the bulbs are connected in parallel, loosening one's connection to the electric circuit would increase the brightness of the other bulb.* (2Y-5A)

Table 10

The Qualities of the Argument Components in the Group Discussions

		Groups											
		1		2		3		4		5		Total	
		2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y	2Y	4Y
Categories		%	%	%	%	%	%	%	%	%	%	%	%
Claim	Right	72	93	47	48	78	84	83	93	20	85	64	82
	Wrong	22	7	2	38	5	0	11	0	80	5	22	11
	Misconception	4	0	27	14	17	6	6	7	0	0	14	7
Warrant	Sufficient	0	0	0	0	0	0	0	11	0	0	0	3
	Insufficient	90	100	31	40	56	83	100	74	71	82	64	75
	Misconception	10	0	68	60	44	17	0	15	28	8	36	22
Data	Sufficient	0	0	0	0	0	0	0	17	0	17	0	6
	Insufficient	100	100	66	88	100	67	0	83	100	50	92	74
	Misconception	0	0	33	12	0	33	0	0	0	33	8	10

This decreased to 5% at the end year four. The reverse situation was observed in the claims produced among the second group of participants. While only 2% of the claims presented by the second group were wrong at the end of year two, this percentage rose up to 38% at the end of year four. The claims having the highest number of misconceptions were the second group discussing free fall. The claims having misconceptions decreased in general at the end of year four, and no claims were detected to have misconceptions in the fifth group's discussion.

When claims were analyzed in terms of being sufficient, insufficient, and having misconceptions, no warrants that addressed and supported the claim from many aspects were identified in any of the group discussions made at the end of year two. There were participants who supported their claims from only one aspect. Only in the fourth group's discussion was there sufficient warrants at the end of year four:

> *I begin first. Water, pure water boils at 100 degrees centigrade under atmospheric pressure. The characteristics of water, its purity, and its location and therefore its atmospheric pressure, they all change the boiling point. So, the differences in measurements may be due to these. The purity of water may be disrupted, changed or the atmospheric pressure where the boiling point is being measured may be increased or decreased. Thus, for the option A, I believe the different measurements may be the consequences of these changes.* (4Y-4E)

However, the rate of the warrants having misconceptions did not decrease in this group. It was found that the arguments examined were also of insufficient quality like the warrants. The fourth group developed arguments with no data at all at the end of year two. When looking at it from this viewpoint, the groups, however, produced a higher number of arguments using more data at the end of year four. Although misconceptions were not embedded in the data that much, misconceptions were identified to be present in the second group's discussion at the end of both years and in the third group at the end of year four. The data deemed to be sufficient were observed in the fourth and fifth groups' discussions at the end of year four.

Understanding the Nature of Scientific Argumentation: Competing Theories

When the arguments presented in the discussion on conflicting theories were examined in terms of quality, the rate of scientifically wrong claims decreased at the end of year four and, although some claims had misconceptions in the discussions conducted at the end of year two, there were no claims with misconceptions at the end of year four. The wrong claims were usually constructed when option F in the question was discussed:

> *I think it would deteriorate because the air flows through here, the s pipe, that is.* (2Y-1F)

At the end of year two, there was only one participant from this group who presented a moderate argument which included claim, data, and warrant; however, the claim made for the argument was not scientifically right and its warrant was weak:

> *There is a matter that serves as an insulating compound; therefore, as the soup will not contact with air, it can stay as it is forever.* (2Y-1D)

The warrants having misconceptions were present at the end of year two, but there were none detected at the end of year four. The warrants were usually based on the prevention of air contact. This warrant was presented on an activity sheet and the participants did not produce any other at the end of years two and four:

> *It can endure for a few days since it has no contact with air.* (4Y-1D)

The only different warrant presented was found to be the one that was produced at the end of year two, but with misconceptions:

> *The speed of putrefaction decreases related to working of enzymes.* (2Y-1A)

When the discussion process was analyzed, the other issue that caught attention was that none of the participants was able to realize that the first theory given for the question had misconceptions and the other theory was scientifically valid. Instead, all of their focus was centered upon matching the theories to their statements. Also, all the data used in this group were weak.

Constructing an Argument: Free Fall

When the arguments presented in the discussion on free fall were examined in terms of quality, the rate of scientifically right claims did change at the end of years two and four, the rate of wrong claims increased at the end of year four, and the rate of claims with misconceptions decreased. It was identified that the scientifically wrong claims were mostly present in options A and C of the question. However, the claim and its warrant were scientifically right, although the rebuttal presented for the claim had misconceptions:

> *Of course, this is only true if the volume is big, because then the frictional force would apply more, so while the speed of the brick would decrease, the coin would speed up.* (2Y-2C)

The two strong arguments, identified while the argument elements were being examined, were produced by the same participant. The warrants this participant presented for the arguments were for option A of the question and explained the claim from only one aspect, but had misconceptions:

The volume of brick is bigger. Of course, if it is a full brick. As the volume of the brick is bigger, more frictional force applies. (2Y-2C)

When the warrants of the arguments were analyzed, no sufficient warrants were identified supporting the argument from different aspects. Although the rate of claims with misconceptions was high, the misconceptions were usually detected in the warrants. The misconceptions emerged in the warrants when the participants were explaining free fall with the help of using the concepts of volume, density, and weight:

As the density of air is close to that of the paper, the fall would slow down at this point. (2Y-2D)

I threw the pencil. It fell due to frictional force. I think there is no frictional force in the air. (4Y-2C)

It was concluded that in addition to the misconceptions on free fall, the participants also had serious misconceptions with regard to the aforementioned concepts. At the same time, the warrants were weak and supported the claims from only one aspect:

It falls. It affects the volume. (2Y-2A)

The data used by the participants to construct their arguments were scientifically insufficient and some also had misconceptions:

I threw the pencil. It fell due to frictional force. I think there is no frictional force in the air. (2Y-2A)

Understanding an Argument: Structure of Particulate Matter

When the arguments presented in the discussion on structure of particulate matter were examined in terms of quality, the majority of claims were scientifically right. The rate of claims deemed to be right increased at the end of year four, whereas the claims having misconceptions decreased in number at the end of year four.

As for the warrants, it was observed that while the rates of warrants with misconceptions and insufficient warrants were very close, the misconceptions detected in the warrants decreased at the end of year four as they decreased in the claims as well. The participants used warrants or rebuttals in a very similar structure. The warrants were expressed by stating a phrase such as, "due to particulate structure" and rebuttals such as, "if there was no particulate structure":

… but if they did not have particulate structure, they could not split. (4Y-3F)

We are able to split it (paper) so it should have many particles. (2Y-3C)

It was this group discussion where the participants produced the arguments by using the highest amount of data. At the end of year two, however, the data used was insufficient, as was the case at the end of year four. Misconceptions were present only when the concept of molecules was mentioned in the data:

> … *These particles should be flexible or should shrink.* (4Y-3F)

At the end of year two, the participants used insufficient data when providing their arguments, whereas at the end of year four, the data also had misconceptions. The data were in general insufficient at the end of both years.

Interpreting Empirical Data: Boiling Point of Water

When the arguments presented in the discussion on the boiling point of water were examined in terms of quality, the rate of scientifically correct claims was high at the end of both years. While the rate of claims deemed scientifically wrong was 11%, no claims were found to be present in this category at the end of year four. No significant change was observed in the rate of claims with misconceptions. It was found that the wrong arguments were produced when the participants proposed how they obtained the data indicating the boiling point as 105 and 108 degrees:

> *I think it is close to the sea level.* (2Y-4Y)

> *It can be above the sea level.* (4Y-4A)

When the warrants of the arguments were analyzed, they were found to be insufficient in general; however, at the end of year four, this group was the first to provide sufficient warrants that supported the claim from several aspects:

> *I begin first. Water, pure water boils at 100 degrees centigrade under atmospheric pressure. The characteristics of water, its purity, and its location and therefore its atmospheric pressure, they all change the boiling point. So, the differences in measurements may be due to these. The purity of water may be disrupted, changed or the atmospheric pressure where the boiling point is being measured may be increased or decreased. Thus, for option A, I believe the different measurements may be the consequences of these changes.* (4Y-4E)

While there were no warrants that had misconceptions at the end of year two, they were present at the end of year four. In this group, only the fourth year students used data and, except for one, all the others were insufficient.

Predicting and Explaining: Parallel Electric Circuit

When the arguments presented in the discussion on parallel electric circuits were examined in terms of quality, the group at the end of year two supported the highest number of wrong claims; however, this changed at the end of year four and 85% of all the claims were scientifically right.

When the warrants were analyzed, while no sufficient warrants were used at the end of years two and four, the rate of warrants having misconceptions present at the end of year two dropped down at the end of year four. After having examined the arguments comprising claims and warrants produced throughout the group discussion, it was concluded that the warrants were right yet they were insufficient and thus the participants could not provide full explanations and mostly used wrong claims:

> *They are both identical (bulbs) and the same current flows through both, if bulb A's connection is loosened, then more current will flow through bulb B.* (2Y-5A)

The same misconceptions were also found in the warrants and these misconceptions in turn caused the claim to be wrong:

As more current will flow through bulb B, it will be brighter. (2Y-5F)

These data used for the claims were insufficient. When the arguments were analyzed in terms of their scientific quality, it was observed that the participants' misconception was that they used the concepts of current and energy as the same. Only at the end of year four, use of sufficient data was identified.

Conclusion and Discussion

This study investigates the preservice elementary teachers' ability to use the components of an argument structure on given science topics as well as the quality of the developed argument contents at their sophomore and senior. In doing so, it was also aimed to discuss the effect of teacher education programs on the development of argumentation skills of students from years two and four. In this section, the results are discussed in the line of research questions.

The Components of Argument Structures

Toulmin's argumentation pattern was used for the determination of argument structure of the scientific discussions. According to this pattern, an argument's basic components are claim, warrant, and data. However, in more complex arguments, one can encounter qualifiers, backings, and rebuttals as well (Driver et al., 2000). As a result of the discussions analyzed in line with this model, it was found that the arguments proposed by the participants at the end of year two were in general composed of only claim or claim and warrants and were either very weak or weak. Whereas, at the end of year four, it was observed that the participants additionally used data in their arguments which involved claim and warrants and that these arguments were at moderate level. The participants attached particular importance to the use of claims when developing their arguments. When the study conducted by Jimenez-Aleixandre et al. (2000) with high school students was examined, it was seen that the results were similar to those obtained in this study. The students usually focused on the use of claims in their arguments on genetics and instead gave less weight to warrants or evidence to support the claim. Some preservice teachers who participated in the present study (i.e., 1E, 2D, 4B) provided only claims and they used no warrants or data at all. And yet, an argument that lacks warrants and is constructed only around a claim is not valid in general terms (Kaya ve Kilic, 2008). The other preservice teachers, however, did usually use warrants to support their claims. Proposal of numerous claims or claims and warrants was an indication that the participants did not have any problems in producing and supporting their own ideas. These results show similarities to some of the research conducted in the related literature (Aslan, 2014; Cinici et al., 2014; Demirci, 2008; Jimenez-Aleixandre & Rodriguez 2000; Kaya, 2012; Kutluca et al., 2014; Osborne et al., 2004). According to Demicioglu and Ucar (2014), the greater the quality of knowledge on a topic, fewer total number of claims are produced. In this study, too, the results were similar. While the rate of claims produced was as high as 40% at the end of year two, this rate decreased to 28% at the end of year four. Thus, it can be deducted that the content knowledge of those participants who only presented claims on the related subject was not as sufficient so as to help them develop quality and complex arguments (Aslan, 2014).

It is essential to make use of data and convincing evidence to support the validity of a claim when constructing a scientific argument (Clark & Sampson, 2008; Karisan & Yilmaz-Tuzun, 2012; Yildirir & Nakiboglu, 2014). Though increased, the research shows that the rate of arguments produced during all the discussions with data in addition to claims was still quite low (23%). This shows that the preservice teachers were unable to fully correlate claim and data which should be at the core of an argument, as they proposed warrants for their claims without using enough data or evidence. This is because the warrants are in fact the correlations established between the claim and the data (Kaya et al., 2014). In this case, we can argue that the students may actually have inadequate content knowledge on the subject at hand as they only made a claim or did not benefit from data and evidence to establish the claim-data relation.

When Demircioglu and Ucar (2014) examined the relationship between content knowledge and use of data, they found that the total amount of data used increased at the end of the posttests. The researchers arrived at the conclusion that the students were able to enhance their content knowledge throughout the study process and thus could create claims with more scientific reasoning and evidence. According to Sampson and Clark (2011), the students possessing a higher level of competencies in terms of scientific knowledge can foster the generation of more quality and complex arguments. This research, too, yielded similar results. In the arguments constructed at the end of year four, there were more data generated. When this is taken into account from the viewpoint of preservice elementary teachers, although they had completed the sciences courses at the end of year two, the science pedagogy helped them make up their inadequacies in content knowledge.

Another element to promote the quality of an argument for which warrant is made is rebuttals (Driver, et al., 2000; Kaya & Kilic, 2008; Osborne et al., 2004). The rate of arguments that included rebuttals among the statements deemed as arguments in the research was evenly low (11%) at the end of both years. Of the total number of arguments, only 2% of arguments had claims, warrants, and rebuttals and in only 1% of the arguments was data used in addition to said components. At the end of year four, this rate slightly changed, with the rate of arguments constructed with claims, warrants, and rebuttals being 4% and the arguments supported also with data being 1%. The rebuttals were usually presented as counter-claims. The highest number of rebuttals was produced in the second group. Demircioglu and Ucar (2014) claim that the increased number of rebuttals in the construction of arguments and that this number being higher than the warrants and data are an indication that these students are good at refuting ideas and thus the quality of arguments increases. Despite their conclusion, when the discussions of the second group at the end of years two and four were analyzed, it was determined that the rate of very weak and weak arguments was very high (42%) at the end of year two, and the rate of weak (27%) and moderate (47%) arguments was higher than the rate of arguments in other categories. Besides, no participant in the group was able to construct a strong argument at the end of year four. Therefore, the high number of arguments do not signify that the presented arguments are strong (Maloney & Simon, 2006).

The difference between the numbers of arguments among groups lies primarily in the structure of the questions. Yet, this difference can also present itself among groups discussing the same question (Maloney & Simon, 2006). According to Demircioglu and Ucar (2014) proposing more than one claim on a subject may arise due to some participants' lack of knowledge on that specific subject. When looking at the intergroup argument construction skills, each group's number and quality of arguments differ. In some of the groups, participants expressed solely their own opinions, not paying attention to the validity or inaccuracy of the opinions given by others and they did not provide rebuttals for counter-claims, all of which in turn reduced the quality of arguments developed during the discussion process. However, the fifth group discussing electrical circuit, for instance, where a smaller number of arguments was produced, half of the arguments were at moderate level in general at the end of year two. This result supports the findings of the research carried out by Maloney and Simon (2006).

From the aforementioned results, it can be said that the argument construction skills of the participants engaged in the research were weak. The participants' construction of very weak and weak arguments could be related to various reasons. One of them could be that they had never been given argumentation-based training. It has been shown by a great deal of research that the process of developing a valid and strong argument is not a spontaneous one, rather it is a skill that you acquire through education and practice (Hogan & Maglienti, 2001; Kuhn, 1991). Aslan (2014), too, notes in his study that one of the reasons for the failure of students in constructing arguments is that they never had any experience in argument construction and argumentation-based trainings. The findings of the research note that the quality of arguments created by the participants who attended an argumentation-based training increased (Demircioglu and Ucar, 2010; Karisan & Topcu, 2016; Karisan & Yilmaz-Tuzun, 2012; Kececi et al., 2011; Kingir et al., 2010; Yesildag et al., 2010).

Another reason could be the lack of knowledge of the participants on the discussion subjects. As the preservice teachers did not make use of scientific terminology whilst in discussion, this, in particular, can be an indication of their insufficient knowledge. In the related literature, studies support this idea (Aslan, 2014; Sampson & Clark, 2011; Tavares et al., 2010; vonAufscnaiter et al., 2008) and the idea that content knowledge does not directly affect the quality of an argument (Eskin & Bekiroglu, 2009; Khun, 1991; Kutluca et al., 2014; Perkins, et al., 1991). In their study conducted with physics preservice teachers, Hakyolu and Ogan-Bekiroglu (2011) assert that the quality of scientific knowledge does not directly link to overall quality of argument. In their research, Clark and Sampson (2008) are of the view that these findings should also be evaluated in terms of discussion frequency, presence of different opinions of the participants, and their skills of conveying their previous knowledge, rather than just focusing on sufficient scientific knowledge.

The Quality of Arguments

Another objective of the research was to determine the quality of the argument components. To this end, the components were examined in such categories as scientifically right, wrong, sufficient, insufficient, or having misconceptions. When the arguments were analyzed, it can be seen that while the participants were successful at the selection of a claim in general at the end of years two and four, they used warrants that supported their claims from a single aspect. Overall, the groups were able to create correct claims. However, some of these claims had misconceptions embedded in them. Among the groups, except for the fifth group at the end of year two, the presence of wrong claims was rare. When the warrants and data were examined, those that supported the claim from many aspects were not identified. In all the groups, the participants presented only one warrant whilst supporting their claims. These warrants were mostly devoid of scientific terminology. The data used was insufficient. This result was also similar to that in the study conducted by Aslan (2014) with high school students. Aslan (2014) asserts that this result is rooted in the inability of students to make correlations between other concepts as they lacked adequate content knowledge. Similar reasons were found in this study as well. For example, it was observed that the participants did not really include concepts apart from the ones already given in the activity sheet handed out during the discussions. The concepts they used had misconceptions in general and most of the time were not related to the subject in question.

The participants of the second group proposed the highest number of claims with misconceptions and this also applied for the warrants. And this had an impact on the scientific validity of their arguments. It was observed that the participants were unable to fully explain what they meant by the concepts of weight, mass, frictional force, and gravity. The same misconceptions were also detected in the study done by Kocakulah and Kenar Acil (2011) with eighth graders. The participants in the fifth group had the highest number of wrong claims as opposed to other groups. The reason as to why is assumed to be their inability to distinguish the differences between the concepts regarding electrical circuits. As the discussion ended within a very short period of time and the participants only presented their opinions and did not further discuss them, there was not much data. The concepts such as current, voltage, and energy identified in the warrants were used as if they bear the same meaning. These results show similarities with that of Kucukkozer's (2003) research administered to high school students.

The Contribution of The Teacher Education Programs

In Turkey, teacher training program courses did not directly support argumentation-based education. On the other hand, preservice teachers are required to engage in scientific discussions as a part of their courses to enrich their knowledge in science and the teaching of science. Since the preservice teachers participated in the science courses for the first two years and that they followed up the Science Laboratory Practices I and II during the academic year when the study was under way, it was assumed that their content knowledge would be enough to address the discussion subjects at primary school level. Whereas Means and Voss (1996) argue that sufficient scientific knowledge does not influence all the aspects of an argument as much as it affects the backings in particular. As for this study, although the content knowledge which the preservice teachers had was sufficient at the end of year two, it was also deducted that the enhancement of their arguments' quality at the end of year four showed that they were better at comprehending and transferring this knowledge after receiving technology and pedagogy courses where they used their knowledge in practice. The practices embedded in these courses enabled the preservice teachers to present many scientific concepts they had learned previously in theory. It can be argued that these preparations helped them have command of subjects within science teaching.

However, the difficulties they encounter in forming arguments can be attributed to their not having the opportunity to form enough arguments in the courses they had taken during their education program. In the research conducted by Aydemir et al. (2018) with science teachers, they found that preservice science teachers also formed insufficient arguments. The preservice teachers stated that traditional teaching was used intensively in undergraduate courses, the subjects were generally explained verbally, and the argumentation approach was used very rarely in the courses; therefore, they found themselves insufficient prepared in the argumentation approach and argumentation. In support of this view, Sahin et al. (2015) stated in their study with faculty of education that the academics who run science classes carry the argumentation method to their classes, but they often have difficulty due to the crowded classes and the content of the course.

The science laboratory I and II courses, which have a great potential for creating arguments, are designed as a course in which teacher candidates do their laboratory applications. The studies related to these courses reveal that argumentation and inquiry-based learning practice processes contribute to laboratory applications by helping in structuring scientific knowledge (Hohenshell & Hand, 2006). However, research reveals that preservice elementary teachers do not frequently encounter argumentation-based practices during their education (Cappellaro, 2016). Cappellaro (2016) also found that half of the participants who completed the Physics, Chemistry, Biology, Science Laboratories I and II, and Science Teaching I courses stated that they did not take a course that used argumentation-based teaching. Preservice teachers also think that a scientific classroom environment can be created using direct instruction and experimentation. They also mentioned the laboratory as the primary method that can be used in science where argumentation was mentioned only by one of the 36 preservice teachers who participated in the study (Karaer et al., 2020). Among the reasons why these preservice teachers who participated in this study could not produce persuasive and qualified arguments in the discussions in real terms was that they had not received any training aimed at developing their argumentation skills and therefore failed to fully participate in such discussion activity. The second reason could be that the learning settings where preservice teachers can transfer and question the knowledge they acquired, where their awareness can be raised to realize the misconceptions that they or their students' arguments may have, and where scientific issues are addressed and discussed are not provided as much as they should be. From this aspect, it can be concluded that providing preservice teachers with settings where they can produce more arguments during the courses they take and where they can engage in discussions would build their capacity for constructing arguments and help them perceive their misconceptions (Kingir et al., 2011).

Considering the critical role of elementary teachers in preparing students to think, explain, ask, and argue about science and interconnected concepts, science teachers at elementary level need more specific attention considering the variety of courses they must teach. This study contributes to an understanding of preservice elementary teachers' argumentation skills and the role of teacher education programs. According to the results, developing preservice teachers' understanding and experiences in teaching science through argumentation is necessary and intervention programs as a part of science courses also need to be considered. Further research needs to deal with instructional and practical issues.

The authors received no financial support for the research, authorship, and/or publication of this manuscript.

Evren Cappellaro (evrenne@gmail.com) is a lecturer in the department of elementary education at Akdeniz University where she teaches science education and environment education courses to teacher candidates. Her research focuses on scientific thinking, the nature of science, and environmental issues with a particular focus on water-related environmental science.

Demet Seban (demet.seban@alanya.edu.tr) is an associate professor in the Department of Elementary Education at Alanya Alaaddin Keykubat University. Her interest areas include curriculum and instruction, teacher education, and teaching literacy. More specifically, she works on preparing teacher candidates to teach reading and writing, content area literacies, and developing language education programs at the elementary level.

References

Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., Tillmanns, M., & Gilbert, L. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction, 19*(1), 1–46.

Aslan, S. (2014). Ogrencilerin yazili bilimsel arguman olusturma ve degerlendirme becerilerinin incelenmesi [Analysis of students' written scientific argument generate and evaluation skills]. *Journal of Theory and Practice in Education, 10*(1), 41-74.

Aydemir, S., Karakaya Cirit, D., Kaya, S., & Azger, C. (2018). Fen bilgisi ogretmen adaylarinin argumantasyona iliskin gorusleri ve argüman kurma becerilerinin arastirilmasi [Exploring the preservice science teachers' views of argumentation and of their argument development skills]. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, *6*(STEMES'18), 131-138. https://doi.org/10.18506/anemon.470577

Cappellaro, E. (2016, May). *Sinif ogretmeni adaylarinin argumantasyon algıları* [Preservice elementary teachers' argumentation perceptions. [Paper presentation]. 15th International Primary Teacher Education Symposium, Mugla.

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in k-12 science contexts. *Review of Educational Research, 80*(3), 336–371. https://doi.org/10.3102/0034654310376953

Cavagnetto, A., Hand, B. M., & Norton-Meier, L. (2010). The nature of elementary student science discourse in the context of the science writing heuristic approach. *International Journal of Science Education, 32*, 427–449. https://doi.org/10.1080/09500690802627277

Cinici, A., Ozden, M., Akgun, A., Herdem, K., Karabiber, H. L., & Deniz, S. M. (2014). Kavram karikaturleriyle desteklenmis argumantasyon temelli uygulamaların etkililiginin incelenmesi [Investigating the effectiveness of argumentation-based activities supported with concept cartoon]. *Adiyaman Universitesi Sosyal Bilimler Enstitusu Dergisi, 7*(18), 571- 596. https://doi.org/10.14520/adyusbd.839

Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*, 293-321. https://doi.org/10.1002/tea.20216

Cresswell, J. W. (2007). *Qualitative inquiry and research design* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Demirci, N. (2008). *Toulmin'in bilimsel tartısma modeli odakli egitimin kimya ogretmen adaylarinin temel kimya konularini anlamalari ve tartisma seviyeleri uzerine etkisi* [The impact of Toulmin's focus on the scientific debate model of chemistry teacher candidates on understanding and discussing basic chemistry subjects]. (Unpublished master's thesis). Gazi Universitesi Egitim Bilimleri Enstitusu, Ankara.

Demircioglu, T., & Ucar, S. (2014). Akkuyu Nukleer Santrali konusunda uretilen yazili argumanlarin incelenmesi [Investigation of written arguments about Akkuyu Nuclear Power Plant]. *Elementary Education Online, 13*(4), 1373-1386. https://doi.org/10.17051/io.2014.31390

Denzin, N. K., & Lincoln, Y. S. (1998). Entering the field of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research theories and issues* (pp. 1-34). Thousand Oaks, CA: Sage Publications.

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

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

Erduran, S., Ozdem, Y., & Park, J-Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998–2014. *International Journal of STEM Education, 2*(1), 1-12. https://doi.org/10.1186/s40594-015-0020-1

Eskin, H., & Bekiroglu, F. O. (2009). Investigation of a pattern between students' engagement in argumentation and their science content knowledge: A case study. *Eurasia Journal of Mathematics, Science and Technology Education, 5*(1), 63-70.

Hakyolu, H., & Ogan-Bekiroglu, F. (2011). Assessment of students' science knowledge levels and their involvement with argumentation. *International Journal for Cross- Disciplinary Subjects in Education, 2*(1), 264-270. https://doi.org/10.20533/ijcdse.2042.6364.2011.0038

Hohenshell, l. M., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed method study. *International Journal of Science Education, 28*(2-3), 261-289.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinning of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching, 38*(6), 663-687. https://doi.org/10.1002/tea.1025

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. https://doi.org/10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Jimenez-Aleixandre M. P., & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education perspectives from classroom-based research*, (pp. 3–28), Netherlands: Springer.

Karaer, G., Karademir, E. ve Tezel, O. (2020). Proje tabanli ogretimin fen laboratuvari dersinde kullaniminin sınıf ögretmeni adaylarının goruslerine gore degerlendirilmesi [Evaluating the use of project based teaching on the science laboratory course according to the views of the classroom teacher candidates]. *Journal of Qualitative Research in Education*, *8*(1), 73-95. https://doi.org/10.14689/issn.21482624.1.8c.1s.4m

Karasar, N. (1995). *Bilimsel arastırma yontemi* [Scientific research method]. Ankara: 3A Arastirma Egitim Danismanlik Ltd. Sti.

Karisan, D., & Topcu, M. S. (2016). Contents exploring the preservice science teachers' written argumentation skills: The global climate change issue. *International Journal of Environmental and Science Education, 11(6),* 1347-1363. https://doi.org/10.12973/ijese.2016

Karisan, D., & Yilmaz-Tuzun, O. (2012, June 27). *Fen ogretmen adaylarinin laboratuvar uygulamalari dersi deney raporlarinin incelenmesi* [Examination of experiment reports of laboratory practice lessons of science teacher candidates]. [Paper presentation]. X. Ulusal Fen Bilimleri ve Matematik Egitimi Kongresi, Nigde.

Kaya, E. (2012, June 27). *Argumantasyona dayali etkinliklerin ogretmen adaylarının kimyasal denge konusunu anlamalarına etkisi* [The influence of the argumentation-based activities on the students' understanding of the chemical equilibrium]. [Paper presentation]. X. Ulusal Fen Bilimleri ve Matematik Egitimi Kongresi, Nigde.

Kaya, E., Cetin, P. S., & Erduran, S. (2014). Iki argumantasyon testinin Turkce'ye uyarlanmasi [Adaptation of two argumentation tests into Turkish]. *Elementary Education Online, 13*(3), 1014-1032.

Kaya, O. N., & Kilic, Z. (2008). Etkin bir fen ogretimi için tartısmaci soylev [Argumentative discourse for the effective teaching of science]. *Ahi Evran Universitesi Kirsehir Egitim Fakultesi Dergisi, 9*(3), 89-100.

Kececi, G., Kirilmazkaya, G., & Kirbag Zengin, F. (2011, May 16). *Ilkogretim ogrencilerinin genetigi degistirilmis organizmalari on-line argumantasyon yontemi ile ogrenmesi.* [Elementary school students' learning about genetically modified organisms with the online scientific argumentation learning program]. [Paper presentation]. 6th International Advanced Technologies, Elazig.

Kingir S., Geban, O., & Gunel M. (2011). Ogrencilerin kimya derslerinde argumantasyon tabanli bilim ogrenme yaklasiminin uygulanmasina iliskin gorusleri [Students' views on the implementation of the argument-based science learning approach in chemistry classes]. *Selcuk Universitesi Ahmet Kelesoglu Egitim Fakultesi Dergisi, 32*, 15-28.

Kingir, S., Geban, O. & Gunel, M. (2010, September 23). *Argumantasyon tabanli bilim ogrenme (ATBO) yaklasiminin 9. sinif ogrencilerinin kimya kavramlarini ogrenmelerine etkisi* [The impact of the argument-based science learning (ABSL) approach to learning chemistry concepts of 9th grade students]. [Paper presentation]. IX. Ulusal Fen Bilimleri ve Matematik Kongresi, Izmir.

Kim, S., & Hand, B. (2015). An analysis of argumentation discourse patterns in elementary teachers' science classroom discussions. *Journal of Science Teacher Education, 26*(3), 221-236. https://doi.org/10.1007/s10972-014-9416-x

Kocakulah, S., & Kenar Acil, Z. (2011). Ilkogretim ogrencilerinin gozuyle "Yercekimi nerededir?" [From the perception of elementary school students: "Where is gravity?"]. *Turk Fen Egitimi Dergisi, 8*(2), 135-151.

Kucukozer, H. (2003). Lise I. sinif ogrencilerinin basit elektrik devreleriyle ilgili olarak kavram yanilgilari [Misconceptions of first year secondary school students' about simple electric circuits]. *Hacettepe Universitesi Egitim Fakultesi Dergisi, 25*, 142-148.

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

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62, 155-178.

Kuhn, D. (2005). *Education for thinking.* Cambridge, MA: Harvard University Press.

Kutluca, A. Y., Cetin, P. S., & Dogan, N. (2014). Effect of content knowledge on scientific argumentation quality: Cloning context. *Necatibey Egitim Fakultesi Elektronik Fen ve Matematik Egitimi Dergisi, 8*(1), 1-30. https://doi.org/10.12973/nefmed.2014.8.1.a1
Lee, M-H., Wu, Y., Tien, T., & Chin-Chung, A. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education, 31*(15), 1999–2020. https://doi.org/10.1080/09500690802314876
Lin, T. C., Lin, T. J., & Tsai, C. C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education, 36*(8), 1346–1372. https://doi.org/10.1080/09500693.2013.864428
Maloney, J., & Simon, S. (2006). Mapping children's discussion of evidence in science to assess collaboration and argumentation. *Science Education, 28*(15), 1817-1841. https://doi.org/10.1080/09500690600855419
McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education, 94*(2), 203-229. https://doi.org/10.1002/sce.20364
Means, L. M., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition and Instruction, 14*(2), 139-178.
M.E.B. (2015). Fen bilimleri dersi (3-8. sınıflar) öğretim program [Science Curriculum (3-8. Grades)]. Ankara: Milli Egitim Bakanligi.
Okada, A., & Buckingham Shum, S. (2008). Evidence-based dialogue maps as a research tool to evaluate the quality of school pupils' scientific argumentation. *International Journal of Research & Method in Education, 31*(3), 291–315. https://doi.org/10.1080/17437270802417184
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. https://doi.org/10.1002/tea.20035
Osborne, J. F., Erduran, S., Simon, S., & Monk, M. (2001). Enhancing the quality of argument in school science. *School Science Review, 82*(301), 63-70.
Ozdem, Y., Cakiroglu, J., Ertepinar, H., & Erduran, S. (2013). The nature of preservice science teachers' argumentation in inquiry-oriented laboratory context. *International Journal of Science Education, 35*(15), 2559–2586. https://doi.org/10.1080/09500693.2011.611835
Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.), *Informal reasoning and education*, (pp. 83-105). Hillsdale: Erlbaum.
Plantin, C. (2005). *L'Argumentation. Histoire, théories et perspectives.* Paris: Presses Universitaires de France.
Sahin, C., Cakmak, N., & Hacimustafaoglu, M. (2015). Akademisyenlerin elestirel dusunme becerilerinin ve elestirel dusunme becerileri hakkindaki goruslerinin incelenmesi [Investigation of academicians' critical thinking skills and ideas about critical thinking skills]. *Inonu Universitesi Egitim Fakultesi Dergisi, 16*(2), 51-66. https://doi.org/10.17679/iuefd.16211687
Sampson, V., & Clark, D. (2011). A comparison of the collaborative scientific argumentation practices of two high and two low performing groups. *Research in Science Education, 41*, 63-97. https://doi.org/10.1007/s11165-009-9146-9
Sampson, V., & Clark, D. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447-472. https://doi.org/10.1002/sce.20276

Schwarz, B. B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentative activity. *The Journal of the Learning Sciences, 12*(2), 219-256. https://doi.org/10.1207/S15327809JLS1202_3

Tavares, M. L., Jimenez-Aleixandre, M. P., & Mortimer, E. F. (2010). Articulation of conceptual knowledge and argumentation practices by high school students in evolution problems. *Science and Education, 19*(6), 573-598. https://doi.org/10.1007/s11191-009-9206-6

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

Tumay, H., & Koseoglu, F. (2011). Kimya ogretmen adaylarinin argumantasyon odakli ogretim konusunda anlayislarinin gelistirilmesi [Improving the understanding of chemistry teacher candidates on argumentation-based teaching]. *Turk Fen Egitimi Dergisi, 8*(3), 105-119.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*(1), 101-131. https://doi.org/10.1002/tea.20213

Walton, D. N. (1996). Argumentation Schemes for Presumptive Reasoning. Mahwah, NJ: Erlbaum.

Walton, D. N., Reed, C., & Macagno, F. (2008) Argumentation Schemes. Cambridge: Cambridge University Press.

Yesildag, F., Gunel, M., Yilmaz, A. (2010, September 23). *Ilkogretim 8. sinif seviyesinde maddenin yapisi ve ozellikleri unitesini ogrenmede argumantasyon tabanli bilim ogrenme (ATBO) yaklasiminin akademik basariya etkis*i [The impact of argumentation based science learning (ABSL) approach on academic achievement in learning the unit of substances and their properties at 8th grade]. [Paper presentation]. IX. Ulusal Fen Bilimleri ve Matematik Kongresi, Izmir.

Yildirir, H. E., & Nakiboglu, C. (2014). Kimya ogretmen ve ogretmen adaylarının derslerinde kullandiklari argumantasyon sureclerinin incelenmesi [Examination of chemistry teachers and preservice teachers' argumentation processes used in their courses]. *Abant Izzet Baysal Universitesi Egitim Fakultesi Dergisi, 14*(2), 124-154. https://doi.org/10.17240/aibuefd.2014.14.2-5000091531