TRAJECTORIES OF CHANGE IN WRITTEN ARGUMENTS: HOW STUDENTS' SCIENTIFIC WRITTEN ARGUMENTS CHANGE THROUGHOUT A SCHOOL YEAR

By

HEBBAH EL-MOSLIMANY

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey,

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Education

Written under the direction of

Clark A. Chinn

And approved by

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE DISSERTATION

Trajectories of change in written arguments: How students' scientific written arguments change throughout a school year

by HEBBAH EL-MOSLIMANY

DISSERTATION DIRECTOR: CLARK A. CHINN

Science education has moved toward engaging students in the practices of science, including modeling, argumentation, and explanation (Krajcik, McNeill, & Reiser, 2008). One way to engage students in these practices is through model-based inquiry learning environments in which they engage in scientific practices including argumentation using evidence (Lehrer, Schauble, & Lucas, 2008). Students develop verbal and written arguments through interpreting and identifying evidence and using them to support or develop a model that explains the evidence. But there are challenges for students: students may not understand certain aspects of argumentation, including argument construction, evaluating claims and evidence, and using evidence to justify a claim (Driver, Newton, & Osborne, 2000; Sandoval & Millwood, 2005), because this practice is different from the practices with which they are familiar (McNeill, 2011; Ryu & Sandoval, 2012).

To address the challenges of learning to argue with models and evidence, it is important to examine the trajectories of change in students' argumentation along multiple dimensions of reasoning. By exploring students' written arguments at different points of instruction and over an extended period of time, one can look to see when change in

students' arguments actually occurred and how this change occurs (Ryu & Sandoval, 2012). To determine the changes in students' written arguments that can occur over multiple months of model-based inquiry instruction, I analyzed students' written argumentation during a six-month implementation of a model-based inquiry curriculum. In my analyses, I focused on how components of students' arguments changed and identified patterns of change in arguments across time for individual students as well as for higher-performing and lower-performing students.

The purpose of the study was to examine the trajectories of change in students' written arguments, by focusing on how students' reasoning emerges, identifying differences in the quality and structure of students' arguments, and if certain competencies appear first before others develop. This study will help to advance our understanding of how student's reasoning emerges and changes over the school year, and how it can be supported in a model-based inquiry classroom.

ACKNOWLEDGEMENT

First, I would like to start off by thanking all of the members of my dissertation committee for your input and guidance. To my academic advisor and committee chair, Clark Chinn, thank for all of your support, guidance, and words of encouragement. This dissertation would not have been possible without your support and feedback on several revisions of my work. To Ravit Golan Duncan, thank you for your feedback during research meetings, and all of the help you have provided me throughout my time in graduate school. To Rebecca Jordan and Ala Samarapungavan, thank you for agreeing to be on my committee and the insightful comments you both provided.

I would to like to express my sincerest gratitude to the PRACCIS team, especially Elizabeth O'Brien, Shannon Gilbert, and Jennifer Mead, for all of the help you provided me throughout my dissertation. I would also like to thank my graduate school colleagues and friends, Roberta, Veronica, Randi, Na'ama, Rachel, Karlyn, Suparna, La Reina, Dessi, Lynne, Lars, Kait, and Jenn, thank you for your support and friendship throughout my time in graduate school. To my family and other friends, thank you for patience, encouragement, love, and support. And finally, and very special thank you to my parents, you both have always stressed the importance of education and encouraged all of your children to follow their own path. Your love and support have helped make this dissertation reach its conclusion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Efforts to reform science education have led to changes in how science is implemented and in the design of curricula, instruction, and assessments that engage learners and provide norms for how to construct, evaluate, and communicate scientific knowledge (Duschl, Schweingruber, & Shouse, 2007; Schwarz et al., 2009). Science education is moving toward engaging students in the practices of science as well as learning how science functions as a discipline (Duschl et al., 2007; Krajcik, McNeill, & Reiser, 2008). One way to reform science education is to design effective model-based inquiry learning environments in which science is viewed as a sophisticated knowledge building practice leading to a deeper base of knowledge (e.g., Duschl et al., 2007; Lehrer & Schauble, 2005; Lehrer, Schauble, & Lucas, 2008). By constructing model-based learning environments in classrooms, students have opportunities to develop scientific practices such as the construction, revision, and refinement of explanatory models to help develop explanations, make predictions, and make sense of theories and data. Providing opportunities to develop scientific practices is a foundation for the development of students' scientific arguments.

Developing scientific arguments is challenging for students. To address challenges in learning to argue I examined trajectories of change in students' argumentation patterns. For this study, I analyzed students' written arguments at different points of instruction to examine how aspects of students' arguments changed, and if students required certain competencies before others developed. In the next sections, I discuss challenges students' face in model-based inquiry learning environments, the challenges of learning to write scientific arguments and how to support students when they face these challenges. I then discuss trajectories for learning to argue.

Challenges in Model-Based Inquiry Learning Environments

Model-based inquiry leads to positive effects in students' learning and development of scientific knowledge, including the development of conceptual understanding and development of reasoning competencies, (e.g., Berland & Reiser, 2009; Passmore & Stewart, 2002; Passmore & Svoboda, 2012), but there are challenges in the design of effective model-based inquiry learning environments. First, models can be complex (Giere, 2004). For example, models incorporate unseen entities that can be used to represent a scientific concept or idea. Second, in model-based inquiry learning environments students can be introduced to more than one model to explain a phenomenon (Pluta, Chinn, & Duncan, 2011). Using more than one model at a time is challenging for students because they have to evaluate multiple models and use evidence to decide which model explains the phenomenon. Third, model-based inquiry provides students with a new level of autonomy (Berland et al., 2015; Hofer & Pintrich, 1997). Students are responsible for figuring out the answers to a question through the evaluation of a model or models and the evidence provided. As students are introduced to each piece of evidence, they require opportunities to evaluate the quality of the evidence and coordinate which pieces of evidence support a particular model. The goal is that students develop a verbal or written argument. This goal is achieved as students interpret each piece of evidence, identify which pieces of evidence are the strongest and use the evidence to support, refine, or develop a model.

Supporting students as they develop scientific arguments can be challenging for teachers because a model-based inquiry learning environment requires teachers to undergo an intense period of learning new pedagogy, and students need to adapt to new teaching styles and expectations from teachers (e.g., Geier et al., 2008; Schwarz & White, 2005). Additional challenges for teachers include orchestrating verbal or written arguments, providing students support while preserving students' autonomy to make their own decisions when evaluating models and providing assistance when students have difficulty writing arguments.. For students learning within a model-based inquiry environment requires an increase in discussing the models and evidence because they are trying to understand the scientific content (Lehrer, Schauble, & Lucas, 2008).

Challenges of Learning Written Argumentation

There are particular clusters of challenges surrounding argumentation. Students struggle with being autonomous and being responsible for developing solutions. They may not understand certain aspects of argumentation, including argument construction, evaluating claims and evidence, and using evidence to justify a claim (e.g., Driver, Newton, & Osborne, 2000; Sandoval & Millwood, 2005). Because the practice of argumentation differs from what students engaged in previously during science education (McNeill, 2011; Ryu & Sandoval, 2012), students often struggle with identifying their roles in model-based inquiry learning environments. In one study when students were assigned to check each other's work and given specific roles, they were more responsible and autonomous (Herrenkohl & Guerra, 1998). Another challenge for students is identifying the audience reading their argument. When students think their written argument is for their teacher to read, they are more likely to demonstrate what they have learned about a particular phenomenon (Berland & McNeill, 2010) instead of trying to persuade a classmate. Students are explaining what they know, and not trying to convince someone to support their reasoning.

To help support students as they write arguments, McNeill and colleagues proposed the claims, evidence, and reasoning (CER) argumentation framework (Krajcik & McNeill, 2015; McNeill & Krajcik, 2008a, 2008b; McNeill, Lizotte, Krajcik, & Marx, 2006). CER is an argumentation framework that supports students as they learn to develop scientific arguments. Claims are the conclusions that answer the problem that the driving question (defines the task for the lesson or question students are trying to find the answer to) is addressing. Evidence is the data provided to support or contradict the claim, and reasoning can be the link between the model and evidence. Although CER has been suggested as a scaffold because it reminds students of the components of an argument, students struggle with the components of claims, evidence, and reasoning (Driver, Newton, & Osborne, 2000; Sandoval & Millwood, 2005).

To address the challenges occurring in students' written arguments, we need to expand our focus beyond structure and examine the quality of the argument (Chinn, Duncan, Hung, & Rinehart, 2014). McNeill (2011) found that fifth grade students' views about argumentation and their ability to write an argument changed over the course of the year, but the quality of the arguments at different points of the curriculum did not improve as students continued to write arguments. Initially students' written arguments provided information related to the question. Their writing did not include using evidence to justify a claim because most did not know how to construct a scientific argument. The author suggested that by providing students with an argumentation framework like CER, it allowed students to understand the structure of a science argument, but not the quality of the argument.

By the end of the year, most students' written arguments consisted of a claim being supported with evidence and reasoning, but the quality of the argument varied. Quality depended on students' understanding of the science content and the type of question students were asked to answer (McNeill, 2011). If students were presented with an open-ended question, students struggled with identifying the best or most appropriate evidence to support the claim. As for their science knowledge, students needed to understand the content in order to evaluate the pieces of evidence and construct an accurate claim.

For students to be successful at writing a scientific argument, they need to understand the science content and structure of an argument (McNeill, 2011; McNeill et al., 2006). Students were able to follow the argumentation framework (CER) developed by Krajcik and McNeill (Krajcik & McNeill, 2015; McNeill & Krajcik, 2008a, 2008b), but the framework focuses on argument structure. CER does not exhaust issues of argument quality, and we need to address these issues because it is difficult to write a high quality argument (Chinn et al., 2014). One way to address these issues is to provide scaffolds during instruction that help students examine the quality of models and evidence, examine evidence that varies in quality and examine the relationship between the model and evidence that lead to writing high quality arguments.

Supporting Written Argumentation

To support students' scientific reasoning and the development of their written arguments, curricular scaffolds need to be embedded in the curriculum. For instance, Geier et al. (2008) scaffolded students learning through embedded technology. Scaffolds or supports are needed because model-based inquiry challenges both the student and teacher (Edelson, Gordin, & Pea, 1999; Edelson & Reiser, 2006) and students have difficulty providing high quality written arguments. When students are asked to provide an argument they are thereby provided opportunities to justify their knowledge by using evidence to support their claims.

Certain scaffolds included in model-based inquiry instruction can address students' challenges during written argumentation tasks. Prior research on argumentation found that students align their theories with evidence through the use of scaffolds embedded in the curriculum to support students' evaluation of models and evidence. One example is Sandoval and Reiser's (2004) electronic scaffold, ExplanationConstructor, that helped students organize their ideas and construct arguments. It is an electronic journal that helps students articulate their thinking as they investigate a scientific phenomenon. ExplanationConstructor helped to address the challenge of identifying and selecting pieces of evidence in support of a particular claim. Another example is the development, revision, and use of class criteria lists for model goodness (Pluta, Chinn, & Duncan, 2011). The purpose of the list is to help students evaluate model goodness. Students use the list to support their evaluation of competing models. Another example is the MEL matrix that supports students' evaluation of models and evidence. Students can identify which pieces of evidence support either one of the models, both models, or neither model (Rinehart, Duncan, & Chinn, 2014; Rinehart, Duncan, Chinn, Atkins, & DiBenedetti, 2016). This scaffold supports argumentation because students need to evaluate the relationship between evidence and models, and then use evidence to support their reasoning for picking a specific model. Scaffolds support the growth of students' scientific reasoning and knowledge of science content and the construction of a scientific argument (McNeill et al., 2006).

As students are provided opportunities to develop, revise, refine, and evaluate models throughout a curriculum, can we link change to particular scaffolds in the curriculum? For instance, the CER framework is a particular scaffold that can support students in the critiquing and construction of a scientific argument. CER reminds students that they need to use evidence to support their claim (Krajcik & McNeill, 2015; McNeill et al., 2006; McNeill & Krajcik, 2008a, 2008b) and reasoning provides a bridge between the claim and data. Providing a rubric (claims-evidence-reasoning) supports student's construction of a scientific argument, but other scaffolds are needed for students to also improve the quality of their arguments.

An additional challenge for students during written argumentation tasks is identifying who is the audience reading the argument (Berland & McNeill, 2010). One place to start is focusing on the classroom culture and practices. As teachers introduce model-based inquiry to students, they need to establish practices that support students as the students evaluate claims and evidence. One practice to address these challenges during written argumentation tasks is to cast argumentation as an integral part of classroom culture (Ryu & Sandoval, 2012). Another practice to support written argumentation is choosing science topics that are meaningful for students and present science content that has a social context (Bybee, 1987). By including social issues that are meaningful, students become engaged in higher order thinking skills (e.g., evaluating data, making decisions, problem solving, constructing arguments) that are part of a model-based inquiry curriculum (Zohar & Dori, 2003). Other researchers have suggested the establishment of classroom norms and culture (e.g., Berland & McNeill, 2010; Cobb, Stephan, McClain, & Gravemeijer, 2001; Lehrer & Schauble, 2006). Helpful classroom norms include developing a criteria list for model goodness that the whole class uses to evaluate competing models (Pluta et al., 2011).

As designers of model-based inquiry curriculum, the field needs be aware of the challenges and how to support both teachers and students. For instance, Metz (2004, 2008) suggested that students could learn to use inquiry through a gradual process, so curriculum designers should create a learning environment that supports both teachers and students by providing scaffolds or supports that are integrated with the science topics presented throughout the school year. Designers of model-based inquiry curricula for middle school classrooms have implemented this approach (e.g., Pluta et al., 2011: Rinehart et al., 2016; Schwarz et al., 2009), but the field does not fully understand how students' scientific reasoning changes over the course of instruction. One way to identify change is to evaluate students' written arguments at different points of instruction.

Trajectories of Learning to Argue

To understand how to address the challenges of learning to argue, it is important to examine the trajectories of change in students' argumentation. By exploring students' written arguments at different points of instruction and over a longer period of time, one can identify when and how change in students' arguments occurred. The purpose will be to go beyond students' ability to write structurally sound arguments and explore the quality of their arguments. Driver, Newton, and Osborne (2000) found that students struggle with argument construction and the ability to evaluate a claim. Others found that students have difficulty justifying claims with evidence to support their reasoning (Sadler, 2004;

Sandoval & Millwood, 2005; Sandoval, Sodian, Koerber, & Wong, 2014). By looking at trajectories over time, we can focus on what aspects of students' written arguments are changing throughout the curriculum and if there are any patterns to these changes. For instance, when students include higher quality evidence (evidence that supports one model and/or contradicts the other model) does it lead to an increase in reasoning? It is important to understand trajectories because it may lead to improving instruction and identifying what type of support students and teachers need in a model-based inquiry learning environment. Ryu and Sandoval (2012) found that students over an extended period of time improved their ability to construct and evaluate arguments. Part of the students' improvement occurred because students chose a piece of evidence and were afforded multiple opportunities over a long period of time to justify their reasoning for choosing that evidence.

By starting with a review of previous learning progressions in argumentation (e.g., Berland & McNeill, 2010; Osborne et al., 2016), we can address what is already known about argumentation and what remains to be addressed. Both Berland and McNeill (2010) and Osborne et al. (2016) discussed justification of claims through evidence and the value of critiquing an argument in their learning progressions, but they focused more on argument structure. For students to be successful at writing a scientific argument, they need to understand the science content and structure of an argument (McNeill, 2011; McNeill et al., 2006). We also need to understand how the quality of the argument changes over an extended period of time. Both Berland and McNeill (2010) and Osborne et al. (2016) stated that the quality of the argument is dependent on whether the student understands the content, but is there anything else that students need for their arguments to improve in quality? Both learning progressions explain how the construction of arguments becomes more complex as students use evidence and reasoning to support the claim, but neither progression discusses the value in evaluating competing models provided or evaluating the quality of multiple pieces of evidence.

Ryu and Sandoval (2012) suggested that focusing on a longer period of instruction improved students' capacity to argue. Their focus was mostly on verbal arguments, so it makes sense to see if the same findings occur with sustained written arguments. By examining written arguments at different points of instruction, we can see when and how students' arguments changed going beyond exploring how students' arguments improve from pretest to posttest. For instance, by analyzing arguments at different points of instruction, we can identify what aspects of arguments change and how they change over time, and what supports are needed for change to occur. Sandoval, Sodian, Koerber, and Wong (2014) stated that the ability to coordinate a claim and evidence requires various reasoning practices (e.g., quality of the evidence, evidence strength). They also suggest that there may be certain competencies that need to be in place before others can develop. For instance, for students to be able to critique an argument, they need to identify competing arguments and have a good understanding of the models and evidence (Osborne, 2010; Osborne et al., 2016). From an instructional viewpoint, this suggests starting with opportunities where students evaluate competing models and multiple pieces of evidence, that vary in quality, and as they gather an understanding of arguments, then introduce opportunities to critique.

With trajectories, one can see if certain components of written arguments are in place before others, or if a certain component is needed prior to another component

(Osborne et al., 2016). For instance, if students include a specific component in their written argument, does it lead to them including another component in their written argument? Are there particular contexts where students can do dramatically better from pretest to posttest? For instance, if the content in the pretest is easier than the posttest, students' understanding of the content could affect their ability to understand the models and evidence. Before a student can evaluate a model or any piece of evidence, they need to first understand it. Another context is the classroom culture. In classrooms where students are provided multiples opportunities to engage in authentic science practices, they are able to provide justifications (Ryu & Sandoval, 2012) and learn to revise models as new information is introduced (Lehrer & Schauble, 2006).

How does students' ability to write an argument change? To help address this question, we need to think about students' understanding of the content, models, and evidence, their ability to evaluate competing models and evidence that vary in quality and how do the quality of arguments change over time. McNeill (2011) found that over time students' ability to write a structurally sound argument improved, but the argument quality varied based on students' understanding of science content. Students' understanding of argument structure, science content, and ability to critique competing arguments will lead to developing higher quality arguments that go beyond simply following the correct structure of an argument. The *Next Generation Science Standards* (NGSS) states that the competencies of argument structure and understanding science content should be taught at the same time (NGSS Lead States, 2013), and in *Promoting Reasoning and Conceptual Change in Science* (PRACCIS) we introduce students to model goodness and evaluating claims and evidence through using science content (Pluta et al., 2011; Rinehart et al., 2014; Rinehart et al., 2016).

Problem Statement

The purpose of this study is to examine how students' written argumentation changes during the implementation of a model-based inquiry curriculum. Model-based instruction is the focus of prior research because of the important role that models (development, evaluation, revision, and refinement) play in scientific practice. Although there are multiple research studies that focus on model-based instruction, it is still challenging for teachers to implement and students' to participate in. Supports embedded into the curriculum have been helpful, but students still find the practice of argumentation challenging. To identify the challenges that students having with writing arguments, this study will address how students' argumentation changes when students are provided with multiple opportunities to write arguments. This study addresses and identifies what changes occur in argument writing by examining students' written arguments before, during, and after instruction. The reasoning for examining the arguments over a sustained period of time is to identify how change occurs in the arguments and if the changes in written arguments include students writing higher quality arguments, how components of written arguments (model choice, evidence usage, evidence conclusions, reasoning, rebuttals) change, if certain competencies need to be in place before others develop, which supports were useful for argument writing and if additional supports are needed.

This research will help address changes in students' reasoning and determine if supports (e.g., model evidence link (MEL) diagram, evidence ratings, criteria lists, argumentation rubric) used during the intervention supported these changes. To identify changes in reasoning, I started by looking at differences in students' written arguments before and after instruction. I explored reasons for why change did or did not occur by examining trajectories of argumentation. Students were purposefully selected by looking to see which students made greater gains versus students who made lesser gains from pretest to posttest. After determining if there were differences between students within the same classroom and teacher, I identified specific artifacts (written work) throughout the lessons to see how they answered the questions and what type of responses students provided. The artifacts focused on answers to model, evidence, and content understanding questions that were included in different lessons and the posttest, as well as arguments students wrote at the end of the lessons and assessments. The reason for reviewing student work at different points was to determine how student reasoning changed during the implementation of a model-based inquiry curriculum.

Students' scores on the pre and post argumentation assessments were used to compare students' written arguments and identify ten pairs of students. Each pair were two students who have the same teacher in the same class period; the only difference is in how they scored from pretest to posttest (one student making greater gains than the other). Artifacts were chosen that explored students' understanding of content, models, and evidence, interpretation of evidence quality, relationship between the model and evidence, and written arguments. These artifacts provided a picture of the changes that occurred in students' written arguments. After the ten pairs of students were chosen, the written work was coded and analyzed. As part of the analyses, I looked at both individual and group differences to identify similarities and differences between groups and students for their written arguments, their understanding of content, evidence, and models, and if certain competencies needed to be in place before others developed.

Research Questions:

1. What are the trajectories of change over time in students' written arguments?
    a. What are the differences in the quality and structure of students' arguments?
2. In written argumentation, do certain competencies need to be in place before others develop? If yes, what are the competencies?

Ch. 2

Literature Review

The purpose of this study is to investigate trajectories of change along multiple dimensions of reasoning by analyzing students' responses to writing prompts presented before, during, and after a model-based inquiry curriculum. In order to understand how aspects of students' reasoning emerge and changes over the course of a school year, I examined the components of students' written arguments. For each component, I addressed these research questions: (1) what are the trajectories of change over time in students' written arguments? (a) What are the differences in the quality and structure of students' arguments? (2) In written argumentation, do certain competencies need to appear before others develop? If yes, what are the competencies? The literature review begins with a discussion of model-based inquiry learning environments. For us to understand trajectories of change, we first need to review model-based inquiry instruction and learning environments to discuss the challenges that arise in these environments and address ways to supports students' development in argumentation. Model-based inquiry instruction is challenging for students because students are introduced to argumentation while evaluating models and evidence, and using evidence to support the model chosen (Berland & Reiser, 2009; Driver et al., 2000; McNeill et al., 2006).

Model-Based Inquiry Learning Environments

Model-based inquiry learning environments allow students to learn by engaging in authentic scientific practices that lead to better understandings of various scientific concepts. Models are representations that illustrate, explain, and predict natural phenomena (Schwarz et al., 2009). Models help students make sense of theories and data and their evaluation can be identified as a central practice for producing scientific reasoning (Schwarz et al., 2009; Schwarz & White, 2005). As students evaluate, develop, and revise models, they ask questions and use evidence to support their reasoning. Models influence the types of questions and evidence scientists seek to support for a particular argument (Passmore & Stewart, 2002; Passmore & Svoboda, 2012).

Students participating in model-based inquiry learning environments experience some positive learning outcomes, including a greater understanding of scientific content and learning how to construct and justify scientific knowledge (Lehrer & Schauble, 2006; Schwarz et al., 2009; Schwarz & White, 2005; Stewart, Cartier, & Passmore, 2005). The practice of inquiry supports the development of scientific knowledge (Schwarz & Gwekwerere, 2007) and helps students develop their skills for doing science and learning scientific concepts in an in depth manner (NGSS Lead States, 2013; Sandoval, 2005).

As students develop their knowledge of inquiry and modeling, their learning of science content increases (Schwarz & White, 2005). The design of model-based inquiry learning environments provides students with opportunities to develop scientific knowledge and use this knowledge and epistemic ideas to make sense of what they are learning. Students benefit by acquiring deeper understandings of scientific concepts. As part of the authentic practice of science, it is valuable for students to engage in the process of model revision. The process of model revision is especially beneficial when new pieces of evidence are introduced or when students explain their model to the class and questions and comments posed lead acquisition of new understandings of science concepts (Lehrer & Schauble, 2000). This provides an opportunity for making learning meaningful because students are going through a process of model revision and refinement to expand their knowledge of a specific concept. Students' written and verbal arguments provide this opportunity and may help students as they evaluate the models and evidence.

Model-based Inquiry Instruction and Written Arguments

Model-based inquiry instruction provides a foundation for the development of written arguments. An argument is both a justification and a claim focusing on evaluation of claims of knowledge (Osborne & Patterson, 2011). The model of the argument pattern developed by Toulmin includes essential elements of arguments, which are claims, data, warrants, and backings (Erduran, Simon, & Osborne, 2004; Osborne, Erduran, & Simon, 2004). Toulmin's argumentation model is an analytic lens for examining arguments so that when students construct a verbal or written argument they need appropriate evidence to justify a claim (Berland & Reiser, 2009).

The model is often the solution to the problem that is introduced by the driving question posed during instruction. The driving question is defined as a unit-anchoring phenomenon (Berland et al., 2015) that engages students in authentic scientific investigations. Driving questions specify the purpose of instruction by calling attention to a problem that students are asked to solve through the evaluation of evidence and the construction and refinement of models (Krajcik, Blumenfeld, Marx, & Solloway, 1994; Tali, Krajcik, & Blumenfeld, 2006). In the current study, we posed the driving question: "Why do leaves fall?" This driving question was used during an assessment, and the question was a unit-anchoring phenomenon because students investigated this question while evaluating two competing models and five pieces of evidence that varied in quality. Students chose one of the models presented in the unit and used pieces of evidence to support and justify their reasoning. At the end of the task, students provided a written argument; a claim that required justification (Osborne & Patterson, 2011).

Evaluation of competing models is part of the argumentation process because it allows students to make sense of scientific phenomena (Passmore & Svoboda, 2012). As students evaluate, develop, and revise models, they ask questions and use appropriate evidence to support their reasoning and answer the driving question. When students evaluate competing models, they communicate their own ideas or thinking about a particular scientific phenomenon to others in their class. Communicating their ideas provides students opportunities to justify a claim using specific pieces of evidence (e.g., evidence that supports the model chosen and/or contradicts the other model) to support their reasoning.

Argumentation is an essential process for scientists, and when students are provided opportunities to participate in this process they are learning about the social structure of the scientific community (Passmore & Svoboda, 2012). Students learn that scientists discuss phenomena and evaluate models and evidence to develop more robust understandings of phenomena. Students also learn that scientific knowledge is socially constructed (Driver et al., 2000) because scientists discuss their findings with other members of the scientific community. Students are included in the meaningful practice of learning and develop their understanding of scientific concepts and theories. In order for students to build new understandings of science concepts, they need to participate in activities that build onto their existing knowledge (Edelson, 2001; Edelson & Reiser, 2006; Krajcik et al., 2008). To build new understandings, students require opportunities to evaluate, construct, and revise models based on evidence (Schwarz & White, 2005). However, modeling can be challenging for students.

Challenges in Modeling

Models are challenging to construct because they include unseen entities that represent a scientific concept. Models can also be complex (Giere, 2004), and teaching students how to evaluate, construct, and revise models is quite challenging. Students have to integrate scientific practices with scientific content and/or prior knowledge. They have to be active participants in their learning by developing skills such as information gathering, decision-making and problem solving (Bybee, 1987). When students engage in modeling, their understanding of why they were engaging in this practice can vary. They might have engaged in modeling because they were trying to understand a phenomenon they were studying, or to satisfy the expectation of someone else (Berland et al., 2015).

Schwarz et al. (2009) found that as students continue to learn about models and use them for prediction and explanation, their understanding of scientific explanations improved due to metaknowledge. Metaknowledge is defined as understanding the purpose of scientific models and modeling and is identified as a tool for prediction and explanation (Schwarz et al., 2009). The modeling learning progression developed by Schwarz and colleagues (2009) examined the connection between the modeling practice and underlying knowledge needed to support students during the practice of modeling. By incorporating elements of practice with knowledge, students were able to make predictions about closely related phenomena. Even when students evaluated, compared, revised, and identified what to include in the model they still faced challenges in modeling practices, including needing an authentic reason for building a model. Students viewed the purpose of the model creation process as providing an answer to a science question for their teacher instead of making models to communicate their own thinking or ideas. In this instance, students lacked motivation to revise models. The challenge that students face in modeling practices suggests that the design of model-based inquiry learning environments need to carefully support students and teachers who implement the curriculum.

Design of Model-Based Inquiry Learning Environments

Learning environments and curricula for model-based inquiry need to be carefully designed. Previous research on modeling practices suggests learning about modeling and inquiry through a gradual process, the establishment of norms (Cobb et al., 2001; Lehrer & Schauble, 2006) and the use of scaffolds (e.g., Quintana et al., 2004; Sandoval & Reiser, 2004). In addition, students need to be active participants in their learning, and model-based inquiry environments should support students as they learn higher order thinking skills such as evaluating data, problem solving, and argument construction (Bybee, 1987; Zohar & Dori, 2003). This occurs through choosing topics that focus on societal issues (e.g., air quality, climate change, health, genetically modified foods) that affect students' everyday lives and lead them to feel a sense of responsibility. By including science-technology-society (STS) issues in science education, science learning becomes more meaningful and engaging for students and with greater engagement students more readily learn how to evaluate data, make decisions, and construct arguments (Bingle & Gaskill, 1994).

The design of curricula around modeling practices provides students with the opportunity to learn about the specific subject matter and about how to construct and justify scientific knowledge (Passmore & Stewart, 2002). Students learn to use inquiry through a gradual process. For instance, Schwarz and colleagues (2009) stated that the practice of scientific modeling included four aspects of modeling: (1) students constructed models to explain or predict phenomena, (2) used the models to illustrate, explain and predict phenomena, (3) evaluated different models to account for patterns in phenomena and predicted new phenomena, and (4) revised models to increase their explanatory and predictive power by taking into account new evidence. By allowing students to construct or develop their own models, students were able to evaluate their own understanding and learn that as new evidence was introduced the model needed to be revised. The revision of the model reflects an advancement in students understanding of models.

To assist students in modeling practices, Lehrer and Schauble (2006) suggested specific processes that are necessary for designing a model-based inquiry environment. One process, support of scientific inquiry, involves the establishment of norms in the science classroom. Norms included the construction of a criteria list, routines for meetings that occur in the classroom and rules for classroom discourse. To support students' evolving questions about the functioning of aquatic systems, a class established norms for the quality of the research questions and evidence (Lehrer et al., 2008; Lucas, Broderick, Lehrer, & Bohanan, 2005). The teacher asked students to provide justification and judgments about the quality of pieces of evidence that were then summarized into a class criteria list. The list helped students identify what is needed for a good model when they evaluated, revised, and developed models. Throughout the year, the students used the same class criteria to evaluate the soundness of evidence and during research meetings the criteria were referenced and revised. By participating in modeling practices, and being embedded into this community of practice, students learned that initial models need to be revised based on new information provided during these discussions and subsequent evaluation of evidence. When students were able to construct, evaluate, and revise models as part of their scientific practice, it enhanced their learning (Lehrer & Schauble, 2006).

Passmore and Stewart (2002) found similar results in their study that examined the implementation of a model-based inquiry unit on evolution. The nine-week high school unit provided students with opportunities to engage in the development, elaboration, and application of an explanatory model in evolution (Darwin's model of natural selection). Students began by comparing and contrasting Darwin's model with two non-Darwinian models. After they evaluated the models, they spent the rest of the course applying and extending Darwin's model to explain natural phenomena. This provided students with opportunities to reason about evolutionary concepts, such as variation, and assess each other's ideas and explanations. Through engagement in modeling, students developed a rich understanding of the natural selection model and effectively applied the model to explain changes in populations over time.

Within a model-based inquiry curriculum, students create models to express a theory of a phenomenon, evaluate the created models or models provided based on criteria defined by their class (e.g., accuracy and consistency) and revise models to accommodate new ideas or empirical findings. The Model-Enhanced ThinkerTools (METT) curriculum is an instructional approach that enables middle school students to learn about the process of modeling and nature of models while developing inquiry skills and subject-matter knowledge. The development of students' metamodeling knowledge emphasized learning about the nature and purpose of models and supported their understanding of the nature of science (Schwarz & White, 2005). Students learned that models were useful for testing theories, helping people understand science and predicting phenomena. Because the curriculum explicitly showed students how models of force and motion developed, students were engaged in an inquiry process for learning science content. Students' knowledge of physics improved as a result of developing metamodeling knowledge (Schwarz & White, 2005).

In PRACCIS, *Promoting Reasoning and Conceptual Change in Science*, we have designed a model-based inquiry curriculum for middle school students that addresses students' challenges with modeling practices. The curriculum begins with students being introduced to modeling practices through a unit called Introduction to Models. As part of this 2-day unit, students work in groups and by themselves to identify what criteria are needed for good models. By the end of the unit, students create a criteria list that is used to evaluate models (Pluta et al., 2011). As students continue through the curriculum and learn more about models and evaluating evidence, they have opportunities to revise the class criteria list.

As part of the design of PRACCIS, the two introductory units acclimate students to working in a model-based inquiry environment (Pluta et al., 2011; Rinehart et al., 2014). Students begin by evaluating models and establishing a class criteria list for model goodness. As students continue through the curriculum, different scaffolds are introduced and embedded within instruction to support students' understanding and evaluation of multiple pieces of evidence that vary in quality, evaluating model goodness for two competing models, identifying the relationship (support, contradict, irrelevant) between each piece of evidence and the two models and writing arguments. The introduction of scaffolds is gradual, and after each scaffold is introduced, they continue to be used throughout the curriculum.

To support the challenges students have in model-based inquiry learning environments, Sandoval and Reiser (2004) and Quintana et al. (2004) designed electronic scaffolds. Sandoval and Reiser (2004) developed ExplanationConstructor, an electronic journal where students record their investigations. The electronic journal supported inquiry by helping students articulate their thinking as they investigate a scientific phenomenon. The design of ExplanationConstructor emphasized two criteria for explanations: (1) articulation of coherent, causal accounts, and (2) using data to support causal claims. As students used ExplanationConstructor, it helped them organize ideas and construct explanations by offering specific pieces of evidence to select and opportunities to selected evidence to support claims.

Quintana et al. (2004) designed a scaffolding framework that addressed software tools for science inquiry. The framework includes three science inquiry components: sense making, process management, and articulation and reflection. Scaffolding guidelines and strategies for each of these components were provided. Sense making involved the core processes of scientific inquiry. In sense making, students used evidence and concepts in science to help clarify science phenomena (Berland & Reiser, 2009). Students needed to understand the basic science phenomena and also interpret data to help support their reasoning about the phenomena. This was challenging for students because they were being asked to interpret data that may go against their intuitive beliefs and they lacked domain specific knowledge. Sense making must be made explicit by helping students connect how they think about a problem with scientific concepts that are important in that domain (Quintana et al., 2004). This includes the use of software that provides a scientific visualization system (Edelson et al., 1999) to help students focus on the fine details of scientific concepts. Sense making also includes scaffolds that embed expert guidance to help students learn science content and provides examples of how scientists think about a problem. Having scaffolds in place made scientists' processes for sense making explicit to the students because they saw how scientists approached the problem and used evidence and claims to construct arguments.

Process management is the second component of the scaffolding framework, and its purpose was to manage the processes of scientific inquiry. Quintana et al. (2004) identified three guidelines for process management including: providing structure for complex tasks, embedding expert guidance about scientific practices and the handling of nonsalient tasks to reduce cognitive demands. The first guideline, providing structure for complex tasks, includes suggestions for restricting the complexity of task. For instance, only presenting data that is relevant and manageable for learners is essential. This was seen in the design of various model-based inquiry curricula. The models and evidence provided were not limitless but carefully constructed so students are able to manage the size of the data provided but the data were rich enough for it to be educationally useful.

Articulation and reflection are the third components of the scaffolding framework (Quintana et al., 2004) and they are mutually supportive processes. Inquiry involved the construction and articulation of an argument, and reflection helped students monitor their own understanding. As with the other components of scientific inquiry, students faced challenges. Scaffolds helped to reduce these challenges for articulation and reflection. For example, Toth, Suthers, and Lesgold (2002) found students benefitted from instruction that incorporated representational guidance (evidence mapping) with reflection on inquiry criteria during problem-solving investigations. Through an evidence mapping activity, students evaluated information and formulated an inference between theories and evidence. When a new statement was added to the evidence map, it was not connected to other statements. The software allowed students to place lines between two categories to indicate linkages. Evidence mapping provided students an effective way to evaluate information about science by providing a connection between theories and evidence. Students' use of reflection was an integral component for incorporating evidence mapping into the learning environment and the use of both evidence mapping and reflection helped students learn to think and act scientifically. In classroom environments designed to teach students how to evaluate data in regard to theories, evidence mapping was enhanced through the use of the reflective assessment (Toth, Suthers, & Lesgold, 2002).

Krajcik and McNeill (Krajcik & McNeill, 2015; McNeill & Krajcik, 2008a, 2008b) developed the claims, evidence, and reasoning (CER) framework to support students' argumentation. Claims are statements that answer the problem that the driving question is addressing (Krajcik & McNeill, 2015; McNeill & Krajcik, 2008a, 2008b; McNeill et al., 2006). Krajcik and McNeill (2015) define evidence as scientific data provided to support the claim. Data includes readings, tables, or a database; evidence is the scientific facts or data on which one is basing the claims. Evidence is gathered from first or second-hand sources (e.g., data collected or generated by studies completed by researchers), relevant to the problem and supports students as they use the evidence to support the claim (Krajcik & McNeill, 2015; McNeill et al., 2006; McNeill & Krajcik, 2008a, 2008b). Reasoning is the explanation or justification of why the data are appropriate evidence to support the claim. Reasoning can also be the link between the model and evidence (McNeill et al., 2006).

The scaffolds discussed up to this point were used to support students in model-based inquiry learning environments, but there are challenges for teachers who implement model-based inquiry curricula. Osborne, Erduran, and Simon (2004) worked with teachers to incorporate nine argument-based lessons. The lessons involved discussions to identify valuable and necessary pedagogical strategies to help promote and support students' argumentation skills during science class, try out pedagogical strategies and determine if implementation enhanced teachers' pedagogic argument practice and determine if lessons that follow pedagogical strategies support students' argument leading to enhanced quality in student arguments. Teachers explored and developed the practices of initiating argumentation in the classroom. Throughout the two years of the study, teachers gained insight into how to best support argumentation practices in their classroom through the use of materials provided and pedagogical strategies; including the adoption and adaptation of Toulmin's argumentation pattern. This pattern provided students with a method to help discriminate key features of the arguments' claims, justifications, and rebuttals (Osborne et al., 2004).

Situating modeling practices in classrooms allows students to engage in authentic forms of reasoning and provides students with opportunities to make sense of scientific phenomena through argumentation (Passmore & Svoboda, 2012). Instructional activities that are structured around students' articulation and construction of an argument allow students to provide an explanation to others and clarify their own understanding. Science is a social practice. Introducing students to the social and cultural norms of scientists by providing students the opportunity to justify and explain their reasoning about a scientific phenomenon, can provide insight for students about the social aspects of science. .

Challenges to Instruction During Argumentation Activities

In a carefully designed model-based inquiry learning environment, there are challenges to instruction that need to be addressed because students struggle with modeling and teachers need to understand how to support students during instruction. Driver, Newton, and Osborne (2000) found that students struggle with argument construction and the ability to evaluate a claim, and others have found that students have difficulty justifying claims with evidence to support their reasoning (Sadler, 2004; Sandoval & Millwood, 2005; Sandoval et al., 2014). By evaluating arguments at multiple time points during a school year, we identify what aspects of students' written arguments are changing throughout the year and if there are any patterns to explain these changes. By identifying when and how students' arguments change across time and evaluating the model-based inquiry curriculum, we may be able to identify challenges that occur during instruction and implementation. For instances, what supports embedded in the curriculum are useful or are there parts of arguments that students find difficult and what are some ways to alleviate the difficulty of these parts of the argument?

McNeill (2011) found that 5th grade students initial written arguments simply provided information related to the question. Their writing did not include using evidence to justify a claim because most did not know how to construct a scientific argument.

These initial arguments were similar to an informational text because that is what they were familiar with using in school. By providing students the claim, evidence, reasoning (CER) framework, students learned the structure of science argumentation and their arguments moved from simply informational to structurally accurate. By the end of the year most of the students had written arguments consisting of a claim being supported with evidence and reasoning, but the quality of the argument was not as consistent. The quality of the argument seemed to depend on students' understanding of the science content and the type of question students were asked to answer. If students were presented with an open-ended question, they struggled to identify the best or most appropriate evidence to support the claim. Students needed to understand the content to be able to evaluate the pieces of evidence and construct an accurate claim.

In the design of a model-based inquiry learning environment, scaffolds used during instruction support students in many ways. For instance, when writing a scientific argument, McNeill (2011) provided students with the CER framework to use as a way to structure their arguments (McNeill et al., 2006). The CER framework helped students examine and critique their understanding of scientific concepts (Duschl et al., 2007) and supported the processes of articulation and reflection. Through articulation and reflection, students developed their understanding for why phenomena occur, monitored their understanding of scientific concepts and supported their sense making (Quintana et al., 2004).

Scaffolds provide learners with the appropriate support to engage in a practice that may have otherwise been out of their reach (Davis & Miyake, 2004). For example, to help students with constructing scientific arguments, McNeill (2009) examined how teachers used a chemistry curriculum with the explicit goal of constructing scientific arguments to explain phenomena and justifying claims using evidence and reasoning. The curriculum supported the construction of arguments through the use of the CER framework (McNeill et al., 2006). Although teachers used the same curriculum, their instructional practices to support students in writing arguments varied. The teacher who made the largest modifications to the unit experienced the lowest learning gains among students for scientific argumentation compared to other students. During instruction, her students only wrote one scientific argument (when there were three focal lessons) and she provided limited support. The students in the other classes achieved greater learning gains for scientific arguments because they had multiple opportunities to write arguments and received support during this practice. They were exposed to the curricular scaffolds and in their scientific arguments they were able to justify claims they made using appropriate evidence and reasoning. McNeill (2009) demonstrates that the lack of support may limit students' capacity to evaluate each piece of evidence and understanding that certain pieces of evidence can support a specific model leading to a greater understanding of the scientific phenomena.

In other studies, researchers used different types of curricular scaffolds, such as context-specific versus generic (McNeill & Krajcik, 2009; McNeill et al., 2006) to support students' learning of science content and ability to write an argument to explain phenomena. The generic argument scaffold helped students understand the general framework for a scientific argument and strategies behind constructing an argument. McNeill (2009) used the CER framework (McNeill et al., 2006) as the generic scaffold. Context specific scaffolds were supports provided to students in the form of content knowledge to incorporate into an explanation or hints about the task. The interaction between both scaffolds led to gains in students' ability to write a scientific argument. The context-specific scaffolds led to greater student learning regarding evidence and reasoning, but only when teachers' instructional practices provided the generic scaffold of claims, evidence, and reasoning (CER). There was a synergistic relationship between the generic supports provided by the teacher and context-specific curricular scaffolds.

Identifying what activities and scaffolds lead to greater support alleviates some of the challenges that occur. For instance, students find reflecting on scientific practice a challenging aspect of the model-based inquiry learning environment that requires support (Edelson & Reiser, 2006). Students need to understand the importance of reflection and participate in activities throughout the curriculum that support engagement in reflection. Reflection helps students monitor their own understanding and help make sense of their ideas. A way to support student reflection of ideas throughout a model-based inquiry curriculum is having students work individually and as a group to develop and then use a class criteria list to evaluate the goodness of a model (Chinn, Duncan, Dianovsky, & Rinehart, 2013; Pluta et al., 2011). Pluta, Chinn, and Duncan (2011) reported on students' individual lists of criteria generated as an introductory lesson to introduce students to models and identify students' initial ideas about modeling and model quality. As part of the practice for creating the class criteria lists, students worked individually to create a list, discussed their lists with classmates and then as a group developed a class criteria list. Students used the class criteria list to help evaluate model goodness and as students participated in the model-based inquiry curriculum, they revised the list based on new information. By developing and revising the criteria list, students needed to practice reflection, but they were also embedded into a community of practice. They learned that initial models and their idea of model quality needed reflection and revision based on new information and subsequent evaluation of evidence.

As part of scientific inquiry, scaffolds should support the development of students' argumentation processes by supporting purposes of the argument including sense making, evaluating models and evidence, and reflection of ideas and scientific practices. The scaffolds discussed were designed to support students learning in model-based inquiry classrooms to reduce the challenges that arose. By exploring scaffolds that support aspects of argumentation construction, students were supported when asked to evaluate and revise their arguments and attend to evidence or data that goes against their prior beliefs. It helped students develop a better understanding of science content (McNeill & Krajcik, 2009; McNeill et al., 2006).

Supporting Students as they Develop Written Arguments

Because argumentation is an essential part of science and science education, we need to understand what materials and pedagogical strategies promote argumentation during science education. When students engage in verbal or written arguments they often have difficulty justifying their claims (Berland & McNeill, 2010). Students often struggle with selecting appropriate data to use as evidence, using sufficient evidence in their written arguments (Sandoval & Millwood, 2005) and providing reasoning for why they chose specific pieces of evidence (McNeill et al., 2006). These struggles suggest that the development of scientific argumentation skills take time. To support students as they develop written arguments they need: (a) support to understand argument structure and science content; (b) support to evaluate the quality of each piece of evidence; (c) support to identify relationships between evidence and models; and (d) support to write high quality arguments. We need to identify what processes and strategies can help support and facilitate high quality argumentation in the classroom in order to help teachers during implementation.

Argument Structure

Berland and Reiser (2009) identified three goals for argumentation construction: sense making, articulating and persuasion. Through sense making, students use evidence and concepts in science to help make sense of science phenomena and can provide a good lens for analyzing written arguments. Articulating allows students to explain or state their understanding either in written work or during group discussions. Persuading provides students the opportunity to discuss their argument with peers and use evidence to support their reasoning. Through persuasion, students use the ideas of science to connect evidence to claims.

Berland and Reiser (2009), demonstrate how students use evidence to develop claims, make sense of phenomenon to articulate their understandings, but have difficulty persuading others. Although there are pedagogical strategies and aspects of the argument students can use as a framework to support the development and evaluation of their arguments, it is still challenging for students to engage in argumentation practice. To get a better sense of these challenges and support scientific argumentation in the science classroom, Berland and Reiser (2011) identified two challenges: students rarely revised their ideas when new questions were posed, and students struggled to simultaneously engage in sense making (constructing claims, questioning, and revising) and persuasion (defending an idea). Although students participated in scientific argumentation, the lack of revision of ideas suggests they created an initial claim and did not evaluate new ideas, evidence, scientific concepts or feedback from the teacher or peers. Without the revision of ideas, students are not continuing to develop their scientific knowledge or reasoning.

When students were provided multiple theories and time to reflect, discuss and argue how pieces of evidence did or did not support the theoretical explanation or model, there were improvements in the nature of discourse in the classroom. Some aspects of this intervention were used to support students written arguments, but the process for verbal arguments differed from written arguments. Berland and McNeill (2010) found that students' written arguments were less complex than verbal arguments. In verbal arguments, students used evidence that was appropriate, reasoning, and rebuttals to support claims. In written arguments, they did not include rebuttals; this indicated that their written argument was less sophisticated. The gap between written and verbal arguments may exist because students' abilities to communicate complex argumentative thoughts through written work may not be sufficient. In addition, during verbal arguments, students' interactions with peers provided a reason to develop a rich, convincing argument because they were trying to convince the audience. Whereas with written arguments, students had a different audience and goal - to demonstrate understanding of a phenomenon to the teacher. The written argument could be lacking a key component, such as rebuttal, because students were not considering counterarguments.

The difference between students' written and verbal arguments is a challenge students face. One way to support the development of students' written arguments is to identify the audience students are writing to (Berland & McNeill, 2010). Another challenge for students is the ability to persuade peers. Berland and Reiser (2009) found that students used evidence to justify their claims, but had difficulty with persuading peers. Persuasion is a challenging task for students because they have to connect evidence to a claim. To support students written arguments we need to address these challenges and identify what competencies students need to construct a high quality argument.

Identifying what Competencies Students Need for Argumentation

To understand how students' arguments develop and change over a period of time, we need to review what competencies are part of the argumentation process. McNeill (2011) found that students' ability to construct written arguments changed over the course of the school year, as did their views of scientific explanations, evidence and arguments. McNeill (2011) demonstrates that scientific argumentation was explicitly addressed and taught, which helped students gain a greater understanding of the argumentation practice (e.g., Osborne et al., 2004). At the beginning of the school year, most of the students' written arguments did not include a claim and their writing resembled an informational text, a genre of writing that they had experienced in school before. As part of their instruction, students were provided with the (CER) framework (McNeill & Krajcik, 2008a, 2008b; McNeill et al., 2006), and it supported students in learning about the structure of scientific argument. By the end of the year, most of the students learned the structure of a written argument and wrote arguments with all the structural components, but the quality of their arguments fluctuated (McNeill, 2011). McNeill (2011) and Osborne et al. (2016) both found students' level of scientific content knowledge affected the quality of the scientific argument.

Lehrer, Schauble, and Lucas (2008) suggested that science content and reasoning interact such that the processes of argumentation and the content knowledge of science are developing at the same time. This idea differs from the approach we have taken in PRACCIS where we introduce students to models through an introductory lesson that requires students decide the criteria for model goodness (Pluta et al., 2011) and then incorporates science content into learning what is a good model. Throughout the curriculum, students are asked to evaluate two competing models to identify which model better answers the driving question. Students collaboratively reflect on multiple pieces of evidence, varying in quality, throughout each lesson with the purpose being to evaluate the models and discuss the data as a class, small group, and individually. In addition, group work and class discussions provide students with opportunities to engage in argumentation. The purpose is to provide students support as they learn to develop a reasoned argument that goes beyond following a specific structure (e.g., claims, evidence, reasoning) while using evidence that fits with the model they choose (Chinn et al., 2013; Chinn et al., 2014).

As students engage in scientific reasoning, their conceptual knowledge and inquiry practices need to align (Duschl et al., 2007). This begs the question: should scientific knowledge and inquiry practice be introduced at the same time? Lehrer and colleagues (2008) found that knowledge and the processes of argumentation develop together. However, McNeill (2011) found that when students possessed detailed scientific knowledge of a phenomenon they generated higher quality arguments when following a claim, evidence and reasoning argument structure. Osborne and colleagues (Osborne et al., 2004; Osborne et al., 2016) and Berland and McNeill (2010) agree with McNeill's

(2011) findings and found students' level of scientific content knowledge can affect their ability in constructing a scientific argument.

Further work needs to examine if certain competencies need to be in place before another one develops. For instance, the ability to coordinate a claim and evidence requires various reasoning practices (Sandoval et al., 2014). For reasoning to occur, an argument needs to have both a claim and evidence and then use the evidence to justify the support for the claim. The reasoning practices include understanding the quality of the evidence, strength of the evidence and the evaluation of sources for potential biases. This suggests that certain competencies need to be in place before others can develop.

Learning Progressions for Argumentation

To identify if certain competencies support students' development of written arguments, we needed to understand what progression scientific arguments follow. Do students' written arguments progress gradually or suddenly? Do all students or groups of students follow the same pattern of progression or are there individual similarities and differences? We needed to look across components of written arguments to examine if one competency needs to be in place before another. Previous work on sustained practice of argumentation found that students learned how to apply evidentiary criteria when evaluating and constructing verbal arguments (Ryu & Sandoval, 2012), but this contrasted with previous findings (e.g., Berland & Reiser, 2009; McNeill et al., 2006; Sandoval & Millwood, 2005). We need a better understanding of what progression students follow towards competency in scientific argumentation and mastery of specific scientific content. Examining learning progressions for argumentation can help with determining how scientific argumentation practices develop over time (Berland &

McNeill, 2010; Osborne et al., 2016; Songer & Gotwals, 2012; Songer, Kelcey, & Gotwals, 2009). Learning progressions can include pathways to support student learning, identify key pieces of disciplinary knowledge and practices as increase in levels of complexity over time, and a developmental progression for how scientific understanding develops.

Berland and McNeill (2010) proposed a learning progression for argumentation that focused on pathways to support student learning and the key pieces of disciplinary knowledge and practices that increase in levels of complexity over time. They grounded the progression in research on student learning and science studies of disciplinary practice and explored ways students gradually moved towards engaging in argumentation that aligns with scientific versions of practice. The three dimensions of the learning progression are: instructional context, argumentative product, and argumentative process. Each dimension includes characteristics that change the complexity (moving from simple to complex) of the argument. For instance, the argumentative product increases in complexity as students use more components of the argument. The change in complexity for each dimension provides a way to identify changes in students' argumentation practices and how they can change over time from being simple arguments where students discuss a claim to being able to use sufficient evidence to support their claim.

The first dimension, the instructional context, supports students' engagement in argumentation. This dimension focuses on the complexity of the question presented to students. The question is typically open-ended and may have two or three potential answers. Other aspects of this dimension include the size and appropriateness of the data set and the use of scaffolds. As students move from simple to complex argumentation patterns in this dimension, so does the size and appropriateness of the data presented to them. Students working with simpler data sets include a small data set with limited data, and as students' progress in this dimension they move to a larger data set. Finally at the upper anchor of the dimension, students define their own data set. The appropriateness of the data increases with complexity from students using only data that is appropriate, to evaluating data sets with both appropriate and inappropriate data. The final aspect of this dimension is the use of scaffolds; which range from detailed scaffolds, to moderate scaffolds, to the removal of scaffolds. For the instructional context dimension, as students move from simple to complex, there is a great amount of change occurring in how students engage in argumentation. As students' progress beyond simple questions, data sets, and detailed scaffolds, their understanding of concepts becomes stronger.

The second dimension, argumentative product, focuses on the components of the argument (claims, evidence, reasoning, and rebuttals) and looks at how claims have been justified using evidence. The aspects of this dimension include if students defend claims, the rebuttal of counterclaims and if the components of the argument are appropriate. Initially students can defend a claim, but do not use evidence and/or reasoning to support this claim. As students' understanding of claims move from simple to complex; they move beyond defended a claim, to using evidence to support the claim, and finally defending the claim with evidence and reasoning. Another aspect of this dimension is students' ability to rebut a counterclaim. Students move from not rebutting a counterclaim to rebutting the counterclaim. The final aspect is students' use of evidence, reasoning, and rebuttals appropriately to appropriate and sufficient use. As students' progress on this dimension, there are changes in the argumentative product. The argument moves from identifying a claim (no evidence), to defending a claim with evidence (no reasoning provided), to defending a claim with evidence and reasoning.

The final dimension is the argumentative process. The process examines the interactions between participants in the argument. Argumentative process can be used as a tool to analyze and support students' work in the practice of argumentation and the instructional environment that supports students in this practice. There are two aspects to this dimension. With the first aspect, arguments change in complexity by students either articulating, defending, questioning or evaluating claims (simple), to claims being articulated, defended, questioned, and evaluated (in between simple and complex), and then claims being articulated, defended, questioned, evaluated, and revised (complex). The other aspect examines students' participation in argumentative discourse, moving from teachers prompting students (simple), to students and teachers sharing the responsibility (in between simple and complex), and finally students engaging spontaneously in argumentative discourse (complex). As students move from simple to complex in this dimension their level of understanding arguments and level of comfort increases. The process of argumentation becomes a natural part of their learning environment.

An interesting aspect of the progression is that Berland and McNeill (2010) state that the dimensions are not age dependent. They suggest that students' engagement in scientific argumentation is dependent on their experience with scientific discourse and content. They found that fifth-grade students were able to construct stronger arguments than older students, but this should not occur. If students are provided appropriate support and multiple opportunities to build on their understanding of content and argument structure, students' arguments will become more complex and higher quality. Students' argumentation ability develops over time and the three dimensions (instructional context, argumentative product, and argumentative process) should gradually shift. For instance, for younger students, the instructional context should be simple and include a defined question, detailed scaffolds and provide only relevant data, leading students to write more complex arguments. For older students, the instructional context should be more complex and so should their arguments.

Metz (1995) found that younger elementary school students use evidence to support claims, and Berland and McNeill (2010) found that younger students, in elementary school, might struggle with the reasoning component of scientific argumentation. However, with guidance, fifth-grade students can engage in meaningful argumentation. One suggestion is that students in elementary school are not familiar with providing scientific arguments in school and when students' views about argumentation changed so did their ability to write an argument (McNeill, 2011). Another suggestion is that argument quality depends on students' understanding of science content (Berland & McNeill, 2010; McNeill et al., 2006; Osborne et al., 2004; Osborne et al., 2016). Students with a stronger understanding of content construct stronger arguments.

Osborne et al. (2016) suggested that investigating argumentation through the lens of a learning progression provides a way to identify and test the developmental progression of students' argument and hypothesized a three-tiered learning progression. Their proposed learning progression draws on Toulmin's model for arguments, which is also used by Berland and McNeill (2010). A key distinction between the two argumentation learning progressions is Osborne et al.'s inclusion of critique. The learning progression proposed by Osborne et al. combines construction and critique and explains that when students critique an argument they need to be able to evaluate and discuss competing arguments. This requires analysis to identify the elements of the argument (e.g., claims, warrant, data) and evaluation of the validity of these elements by drawing on conceptual knowledge. As for the construction of an argument, it requires students to remember appropriate information and construct a relationship between a claim and its supporting evidence.

Osborne et al.'s (2016) learning progression consists of three broad levels of argumentation --levels 0, 1, and 2; each level requires the student to have more connections between claims and pieces of evidence. The levels are identified as: zero degrees of coordination (no explicit connection or warrant between claims and evidence), one degree of coordination (students need to make one explicit logical connection between claim and evidence), and two or more degrees of coordination (students compare two or more warrants). This progression differs from Berland and McNeill's (2010) because the categories are broader, it focuses on students' use of evidence to support claims and it distinguishes between students' constructed arguments and critiquing someone else's argument. Another difference is that Osborne et al.'s learning progression is a proposed progression that emphasizes assessing students' arguments, whereas Berland and McNeill focus on instruction. In learning progressions, curriculum and instruction are valuable because it can affect learning trajectories (Duncan & Hmelo-Silver, 2009), and Osborne et al.'s progression is missing a component that emphasizes instruction.

Osborne et al. (2016) discussed how argument quality is dependent on students' knowledge and that students need to be taught argumentation through modeling and instruction (Osborne et al., 2004). However, they did not discuss how argumentation should be taught. They explained that relevant evidence needs to be introduced to students to support argumentation in the classroom, but did not include supports in the learning progression. Whereas, Berland and McNeill (2010) included supports in their learning progression (detailed scaffolds to moderate scaffolds to no scaffolds). By middle school, students can support a claim with a piece of evidence (level 0), provide reasoning for why the evidence supports the claim (level 1), but to reach the highest level (level 2), students need to critique two competing arguments by evaluating and explaining why one argument is stronger than the other argument (Osborne et al., 2016). For middle school students to reach this level they need scaffolding, but Osborne et al.'s learning progression does not discuss what level of scaffolding is needed.

Learning progressions can act as a guide in the design of science curriculum and assessment and offer supports for model-based inquiry practices and students' learning of science content. As students develop more complex understandings of modeling and inquiry, their understanding of science content in specific domains becomes more sophisticated. This leads to the construction of stronger scientific arguments. Further work is needed to understand what aspects of reasoning are captured and how students' written responses within specific science domains change over time. Previous research on argumentation learning progressions found that middle school students were able to identify and make claims, interpret and select valid evidence to support a claim, and provide reasoning that links the claim to the evidence (Berland & McNeill, 2010;

Osborne et al., 2016). Although both learning progressions follow Toulmin's argumentation structure, the progressions differ in what they both address. Osborne et al.'s (2016) three levels of progression are broad, focus on assessing students' arguments not instruction, and include argument critique. Being able to critique an argument is a sophisticated task because students need to be able to evaluate and discuss competing arguments. This requires analysis to identify the elements of the argument (e.g., claims, warrant, data) and evaluation of the validity of these elements by drawing on conceptual knowledge, then creating a counterargument that is relevant to the argument. So students need to have metacognitive knowledge of the nature of the argument and the ability to distinguish the elements. Both learning progressions have gaps that need to be addressed and neither discusses if a specific competency needs to be in place before another one is introduced to students. The focus is on claims, evidence, and reasoning but not in evaluating the quality of these components and how to support students in this process.

What is Missing in the Learning Progressions?

The learning progressions for argumentation provide insight into how change in students' arguments should occur and focus on the foundation of argumentation, and state that a students' level of science content knowledge affects their ability to develop an argument. Both Berland and McNeill (2010) and Osborne et al. (2016) are missing components of argumentation that need to be addressed and added to the argumentation learning progression. Both learning progressions are missing studies where students are asked to discuss criteria for model goodness or criteria for evaluating the quality of multiple pieces of evidence. Another component missing is examining students' ability to evaluate competing models. When students are provided opportunities to evaluate competing models and pieces of evidence that vary in quality; this may lead to a better understanding of counterarguments.

Osborne et al. (2016) suggested that in middle school, curriculum should help students develop a confidence with the notion of claim, evidence, and argument. However, they did not explain if students were provided with claims or evidence and if the evidence varies in quality. The learning progression does not discuss what type of criteria students should use to evaluate the quality of claims and pieces of evidence, and what scaffolds are needed to support students for understanding and evaluating claims and evidence to decide which pieces of evidence are appropriate to support a specific claim. Critique of arguments is an important part of this learning progression, but they do not discuss what type of instruction or support is needed to help students compare and critique arguments.

Berland and McNeill (2010) found that argumentation ability developed over time through students experience in the process of argumentation. With support and extended experience, students' understanding of science content and argument progressed from simple to complex along the three dimensions. When students were initially introduced to argumentation the question posed and the data introduced was limited and amount of scaffolding was high to provide students with an ample amount of support. As students became more comfortable and their understanding of content and arguments progress, questions became more open ended, larger sets of data were introduced and included both appropriate and inappropriate data. Students were able to evaluate the appropriateness of the data and chose the appropriate data to support a specific claim. However, the learning progression does not discuss when students were provided with multiple claims if they were able to evaluate the claims and chose a claim based on which data supported the claim. For students to be able to write a high quality scientific argument they needed to understand the claim, be able to evaluate the quality of each piece of evidence, use the best pieces of evidence to support their reasoning for choosing that claim, rebut counterclaims, and demonstrate that each component (evidence, reasoning, and rebuttal) were appropriate and sufficient.

Berland and McNeill (2010) and Osborne et al.'s (2016) learning progressions both discuss justification of claims through evidence and the value of critiquing an argument. Both are complex processes because students need to understand the content, evaluate the pieces of evidence and claim, and able to discuss competing arguments. However, each progression is missing key components to scientific arguments. There are limitations to these argumentation learning progressions. Both Berland and McNeill (2010) and Osborne et al. (2016) do not address how change occurs throughout a period of time and the best ways to support students for improvement of argumentation. For instance, both learning progressions include students ability to reason, but do not explain how reasoning changes across the levels or dimensions except to state that students are able to defend a claim by selecting evidence that supports the claim. Another limitation is that they do not address changes in understanding what makes a high quality argument. For instance, Berland and McNeill's learning progression discussed the appropriateness of argument components, but does not indicate how students evaluated the quality of competing models using multiple pieces of evidence. Nor do the authors explain what type of support was needed to help students evaluate the models and evidence. Often, evaluating arguments has focused on the structural components of the argument such as claims, evidence, and reasoning (McNeill et al., 2006; McNeill & Krajcik, 2008a, 2008b), but it fails to distinguish “good” arguments from “bad” arguments (Chinn et al., 2014). We need to go beyond focusing on the structure of argument and examine the quality of the argument and what components lead to improvement in argumentation (Chinn & Rinehart, 2016; Chinn, Rinehart, & Buckland, 2014). By going beyond analyzing the structural components of the argument and examining components of arguments to identify improvements in argument quality, we can provide a new insight into argumentation practices.

The purpose of this study is to address some of these limitations and add new insight and ideas to the argumentation learning progressions. By focusing on written arguments at different points of the curriculum, I take a different approach from the argumentation learning progressions because neither examined written arguments at different points of a curriculum, but examined verbal and written arguments for different age groups (Berland & McNeill, 2010) or focused on assessment of middle schoolers (Osborne et al., 2016). Examining arguments at different points of the school year allowed me to identify changes in the quality of the student’s arguments and what components (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals) of students’ arguments improve. The analysis focused on components of the arguments to see if change is gradual or sudden, if similar students’ progress similarly or differently and if certain components of an argument need to be in place before others. For instance, when students provide a reason in their argument, which pieces of evidence do they chose to support their model, and do they discuss why the evidence supports the model? This leads to examining the quality of the argument and supports the idea that for students to produce high quality arguments, epistemic criteria, and processes need to be included in the instruction of argumentative practices.

Theoretical Framework

In order to understand how aspects of students' reasoning emerge and change over the course of a school year, we need to investigate trajectories of change along multiple dimensions of reasoning. The argumentation learning progressions (Berland & McNeill, 2010; Osborne et al., 2016) begin to address how students develop their ability to reason scientifically, but not how it occurs. Both represent a linear movement from students having a simpler understanding to one that is more complex (Duncan & Rivet, 2018), but both need to be refined. Osborne et al.'s (2016) learning progression focused on the levels for argument construction and critique, but does not include what type of instruction or scaffolds are needed for students' arguments to move from a level 0 to a level 2. The three broad levels of argumentation require more connections between claims and evidence, and they do not explain what questions students were asked, how many claims were provided, how they evaluated the claims, what types of evidence is provided or if they have to get their own evidence and if the evidence varies in quality. Berland and McNeill (2010) included claims and evidence in their learning progression and showed how the claims and evidence moved from simple to complex, but neither learning progression addresses students' evaluation of competing models and how it can affect their understanding of counterargument.

Both Berland and McNeill (2010) and Osborne et al. (2016) state that students' understanding of scientific content affected the quality of the argument, but what if students have a similar understanding of content and their arguments differ? We need to understand if there are other aspects of instruction and support that lead to change in argument quality. Prior research on scientific written argumentation found that students' arguments can change over the period of time (McNeill & Krajcik, 2009; McNeill, 2009, 2011), but the focus has often been on the structure of the argument and not on the quality. In one study that addressed argument quality, McNeill (2011) found that by the end of curriculum, students' argument structure improved, but the quality of their arguments fluctuated. Differences in the quality of students' arguments were attributed to students' scientific content knowledge and the scaffolds that were provided for students to answer a particular question. We are still uncertain if specific competencies need to be introduced or taught first, if content knowledge and argumentation should be taught at the same time, and what the focus of instruction in argumentation practices should be. If we want to go beyond structural components of the argument and emphasize the quality of the argument, students need to learn epistemic criteria so they can evaluate products like knowledge, theories and evidence, and the processes by which epistemic products are produced (Chinn et al., 2014).

The research questions posed and the task that has been designed in this study examine the trajectories of change in students' written arguments and focuses on how students' reasoning emerges. I also examine if there are differences in the quality and structure of students' arguments. The final question addresses the issue of competencies and adds to the discussion about whether certain competencies need to addressed first and supported before others develop. This study seeks to further the understanding of how student's reasoning emerges and changes over the school year and how students' scientific argumentation practices can be supported in a model-based inquiry classroom.

Chapter 3

Methods

The purpose for this study is to examine the trajectories of change in students' written arguments to capture how students' reasoning emerges. To determine changes in students' written arguments that occurred over multiple months of model-based inquiry instruction, I selected tasks that asked students to provide a written argument supporting model choice or evaluating a model they had created. Tasks were selected from students who participated in one year of the model-based inquiry curriculum. An example of a choice task is: "Pick which model you think is better. Write an argument to support the model you chose. Write to someone who may not agree with you. Use your argumentation rubric to help write your argument." An example of an explanation task is: "Individually, explain why your model is a good model. Explain which evidence supports the model, and explain in detail how the evidence supports your model." Both of these tasks were designed to elicit students' written arguments, and were similar in format and content to the tasks students completed for the pretest and posttest assessment. By choosing similar tasks, analysis focused on how components of students' arguments changed and identified patterns across arguments for both individual students and two groups of students.

Research Design

Participants

The overall study sample included 284 students in 15, seventh-grade science classrooms, taught by four teachers in a suburban middle school in in the Northeast of the US. Based on the state report card, 50.5% of the students in the school were White, 37.6% Asian, 5.9% Hispanic, 5.2% Black, and 0.9% other. 17% qualified for free or reduced lunch, and the performance of this school was above the state average. The teachers implemented a life-science model-based inquiry curriculum over the course of 20-22 weeks.

Before and after the implementation of the curriculum, students took a pretest and posttest. There were two versions of the assessment (muscle pain and falling leaves) that were designed to be comparable. The two assessments were counterbalanced between the pretest and posttest (i.e., some students received one version as a pretest and the second version as posttest, whereas others completed the assessments in the opposite order). As part of the assessment, students read about two competing models and five pieces of evidence that varied in quality. A driving question defined the task for the assessment and one model was supported by more pieces of evidence than the other model. A description of the curriculum and detailed explanation of the scoring criteria for how participants were chosen is discussed in further detail in this chapter.

Two groups, ten students each, were purposefully selected (Patton, 2002) for this study, from a larger group of seventh-grade students who participated in a model-based inquiry curriculum. The twenty students were selected based on the gains they made from pretest to posttest. Students were selected through a scoring process described in detail below and placed in contrastive pairs. Each pair of students included a student who made greater gains from pretest to posttest versus a student who made lesser gains.

Instructional Context

The PRACCIS team co-designed, in collaboration with teachers, the instructional modules of the life science model-based inquiry curriculum for middle school students.

The instructional materials and supports were designed to prompt students to create, revise and evaluate models based on evidence. Students also engaged in written and verbal argumentation about the models, evidence, and criteria. Topics included natural selection, genetics, and cell organelles. The curriculum involved individual, group, and class activities. A scaffolding suite was included to support students' evaluation of models and multiple pieces of evidence, and their written arguments (Chinn et al., 2013; Chinn, Duschl, Duncan, Buckland, & Pluta, 2008; Pluta et al., 2011; Rinehart et al. 2014; Rinehart et al., 2016).

The scaffolding suite included a model quality criteria list, model-evidence link (MEL) matrix, evidence ratings, and an argumentation rubric introduced at different points of the curriculum. Table 3.1 illustrates the introduction of each scaffold during the curriculum.

Table 3.1

Scaffolds Introduced During the Curriculum

Assessments	Units	Lessons	Introduction of Scaffolds	Length of Time (number of days)
Pretest	(Muscle Pain/Falling Leaves)			One day
	1.Introduction to Modeling		Model Quality Criteria List	Two days (total)
		1. Earthquakes and Reasons		One half Day
		2. Volcano and Model Comparisons		One and one half Days
	2. Sam Spade		MEL matrix	Two Days
	3. Cells			Eight Days (total)
		1. Chloroplasts		Four Days
		2. Nucleus	Evidence Ratings	Four Days

	4. Genetics			Twenty Days (total)
		1. What are Heritable Traits?		One Day
		2. Testing Student Models		Two Days
		3. Strong Gene Model		Three Days
		4. Terminology and Lab		Four Days
		5. Punnett Squares		Three Days
		6. Genetics Jeopardy		One Day
		7. Existence of HIV Resistance	Argumentation Rubric	Three Days
		8. Mechanism of HIV Resistance		Three Days
Modeling Assessment				One Day
	5. Evolution			Thirteen and one half Days
		1. Assessment and Introduction		One Day
		2. Variation and Selection		One Day
		3. Mountain Sheep		Two Days
		4. Moths		One and one half Days
		5. Snakes		Two Days
		6. Generalized Model		One Day
		7. Applying Generalized Model		One Day
		8. Finches		Three Days
Posttest	(Falling Leaves/Muscle Pain)			One Day

Throughout the curriculum students used these various scaffolds to support their reasoning about the relationship between the models and the body of evidence. The first scaffold students were introduced to was the model-quality criteria list. During a two-day unit, Introduction of Modeling, students contrasted a range of scientific models that varied in purpose and quality. Students discussed the models as a class, in small groups and individually to generate lists of criteria for what made good models. Students first developed criteria individually, followed by a class discussion in which they collaboratively developed and agreed on a class list. Among their model evaluation criteria, students brought up issues of evidentiary support, pertinence to the question at hand, clarity (including using graphs and images), appropriate levels of complexity, interest, and others. The generated criteria lists showed that students understood a wide range of modeling ideas and understood a variety of criteria (evidence, accuracy, complexity, and sequence) for model goodness (Pluta et al., 2011). Criteria lists provided a support or scaffold to help students evaluate scientific models introduced throughout the curriculum. As students continued through the units and learned about modeling, they revisited, and revised these criteria lists for evaluating model goodness (Rinehart et al., 2014).

The next unit, Sam Spade, was a two-day lesson focused on introducing the MEL matrix. Figure 3.1 provides an example of the MEL matrix. The MEL was another scaffold used in this curriculum to support students as they evaluated models and evidence and helped them to decide how each piece of evidence either supported, contradicted or was irrelevant to the models the student evaluated. The main purpose of this scaffold was to support students' thinking about the relationship between the model and pieces of evidence (Rinehart et al., 2014). The MEL arrows helped students link each piece of evidence to the models and helped them with selecting which model was better for answering the question or solving the problem. Within the MEL matrix, students also had opportunities to rate the quality of evidence. Students evaluated and then rated the quality of each piece of evidence on a scale from 0-2. Evidence quality ratings were defined for students in the nucleus lesson, one of the lessons in the Cell Unit, as, “when evidence is good, you can believe the conclusions” and “when evidence is bad, you cannot believe the conclusions”. Starting with the nucleus lesson students read and evaluated each piece of evidence and rated how good that piece of evidence was using the numbers 0, 1, and 2. “Rating of 0: This is very bad evidence. This means we cannot believe the conclusions at all. We should ignore this evidence and not think about it anymore. Rating of 1: This is not very good evidence, but it is not totally bad evidence. This means: We can believe the conclusion a little, but we have to be careful. Rating of 2: This is good or very good evidence. This means: We believe the conclusion. The evidence shows that the conclusion is correct.” After reading the evidence statements students individually rated the goodness of each piece of evidence and discussed it in pairs. Students were able to change the evidence ratings after discussing them. Students recorded their rating for each piece of evidence in the rating box. The MEL matrix scaffolded students’ thinking and written arguments by providing space for students to rate evidence and identify relationships between pieces of evidence (Rinehart et al., 2014).

Support	$\rightarrow$
Strongly Support	$\Rightarrow$
Contradict	$\nrightarrow$
Strongly Contradict	$\nRightarrow$
Irrelevant	$\dashrightarrow$

Evidence Quality Rating	Model 1 ATTACK-AND-DESTROY	Model 2 HIV-CAN'T-GET-IN
#1. The Burke Family. 		
#2. How the Body Fights the Hepatitis Virus. 		
#3. Jen's Favorite Genes. 		
#4. CCR5 Research. 		
#5. DNA Analysis 		

Figure 3.1 MEL matrix

The criteria lists, the MEL matrix and evidence ratings, which were introduced early in the curriculum, were incorporated into the rest of the curriculum to support students as they engaged in written and verbal argumentation about the models, evidence and criteria. A final scaffold introduced to students was the evidence-based argumentation writing rubric. Table 3.2 illustrates the rubric introduced in the Genetics Unit to scaffold students' written arguments. The purpose of the rubric was to have students' self-assess the quality of their arguments. For instance, in a good written argument students should use pieces of evidence to support their claim by describing each piece of evidence and how the evidence supported the claim.

Table 3.2

Evidence-based Argumentation Writing Rubric

	0	1	2	3
Number of evidence discussed	No evidence mentioned	Only one evidence discussed	More than one evidence discussed	ALL relevant evidence is discussed
Writing about evidence	No evidence mentioned	Mentions evidence	Mentions and describes evidence	Mentions, describes AND explains how evidence supports the model
Counter evidence	No mention of evidence that contradicts other model	Mentions evidence that contradicts other model	Mentions and describes evidence that contradicts other model	Mentions and describes evidence AND explains that contradicts other model

The suite of scaffolds designed for this curriculum support students in developing high quality written and verbal arguments. In the next section a description of the curriculum and example of a typical lesson are discussed.

Curriculum

Four teachers implemented the life science model-based inquiry curriculum over the course of 20-22 weeks. Before and after the curriculum, students took a pretest and posttest (muscle pain and falling leaves). Table 3.3 provides a timeline for the intervention. After the first two introductory lessons described above, the remaining units—cells, genetics, and evolution—used scaffolds introduced and the evidence ratings and argumentation rubric to support students' written and verbal arguments.

Table 3.3

Timeline of Intervention and Order of Units and Lessons

Assessments	Units	Lessons	Purpose of Lesson	Length of Time (number of days)
Pretest (Muscle Pain/Falling Leaves)				One day
	1.Introduction to Modeling		Students develop arguments based on reasons as they pertain to modeling.	Two days (total)
		1. Earthquakes and Reasons	Students learn to support claims with reasons in a scientific context.	One half Day
		2. Volcano and Model Comparisons	Students discuss and identify what qualities are needed for a good model and develop a model quality criteria list.	One and one half Days
	2. Sam Spade		Students learn to evaluate a body of evidence and two different models by learning about and using the MEL arrows and matrix.	Two Days
	3. Cells			Eight Days (total)
		1. Chloroplasts	Students learn to draw conclusions from various pieces of evidence in order to get an initial idea of the function of chloroplasts in plant cells.	Four Days
		2. Nucleus	Students learn to evaluate a body of evidence and use evidence goodness ratings to evaluate 5 pieces of nucleus evidence. Students determine the best model for the function of the nucleus. Students revise the best model to make it even better based on what they learned from the pieces of evidence.	Four Days
	4. Genetics			Twenty Days (total)

		1. What are Heritable Traits?	Students learn to distinguish between traits that are inherited or acquired, and generate an initial scientific model of inheritance.	One Day
		2. Testing Student Models	Students learn to revise their initial models of inheritance using the model goodness criteria and accounting for three provided pieces of evidence about patterns of inheritance.	Two Days
		3. Strong Gene Model	Students learn to use pedigrees as evidence to generate rules for inheritance, and explain how a trait skips a generation.	Three Days
		4. Terminology and Lab	Students learn to use genetics terminology to describe the physical traits and genes of organisms.	Four Days
		5. Punnett Squares	Students learn to use Punnett Squares to make predictions about likely genotypes and phenotypes of offspring.	Three Days
		6. Genetics Jeopardy	Students learn to distinguish between traits that are inherited or acquired, explain inheritance patterns based on text descriptions genotypes, phenotypes and pedigrees, and use probability rules to infer likely outcomes of different matings.	One Day
		7. Existence of HIV Resistance	Students learn to use scientific evidence to investigate the possibility that genetic resistance to HIV exists.	Three Days
		8. Mechanism of HIV Resistance	Students learn to use scientific evidence to investigate the specific cellular mechanism of genetically based resistance to HIV.	Three Days
Modeling Assessment				One Day
	5. Evolution			Thirteen and one half Days
		1. Assessment and Introduction	Students start to think about how species evolved over time.	One Day

		2. Variation and Selection	Students learn to understand the first step in understanding evolution is learning to recognize variation within species, that variation can be small or large, and the difference between variation that is acquired vs. variation that is inherited.	One Day
		3. Mountain Sheep	Students develop a model for why the horns of mountain sheep have become smaller over the past 25 years.	Two Days
		4. Moths	Students develop a model for why the population of peppered moths has changed color over the past 60 years.	One and one half Days
		5. Snakes	Students develop a model for why the population of red-bellied snakes has changed over the past 75 years.	Two Days
		6. Generalized Model	Students generate a model that explains natural selection based on the previous models from peppered moths and Red-Bellied Black Snakes.	One Day
		7. Applying Generalized Model	Students apply a general natural selection model to particular cases that were introduced to students during Lesson 1.	One Day
		8. Finches	Students develop a model for the change in the finch population after the drought in 1977.	Three Days
Posttest (Muscle Pain/Falling Leaves)				One Day

Pretest and Posttest

Two versions of the pretest and posttest (muscle pain and falling leaves) were designed to be comparable; one about why leaves fall off trees in autumn (FL), and the other about why we feel muscle pain 48-72 hours after exercise (MP). The students were counterbalanced between the pretest and posttest (i.e., some students received MP as a pretest and FL as posttest, whereas others completed the assessments in opposite order). The assessments introduced students to two competing models (Figure 3.2). Students evaluated five pieces of evidence to help them decide which model is accurate. Figures 3.3 and 3.4 illustrate examples of the evidence students evaluated for the FL assessment. After being introduced to the models and evidence, students were prompted to answer: "Which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer."

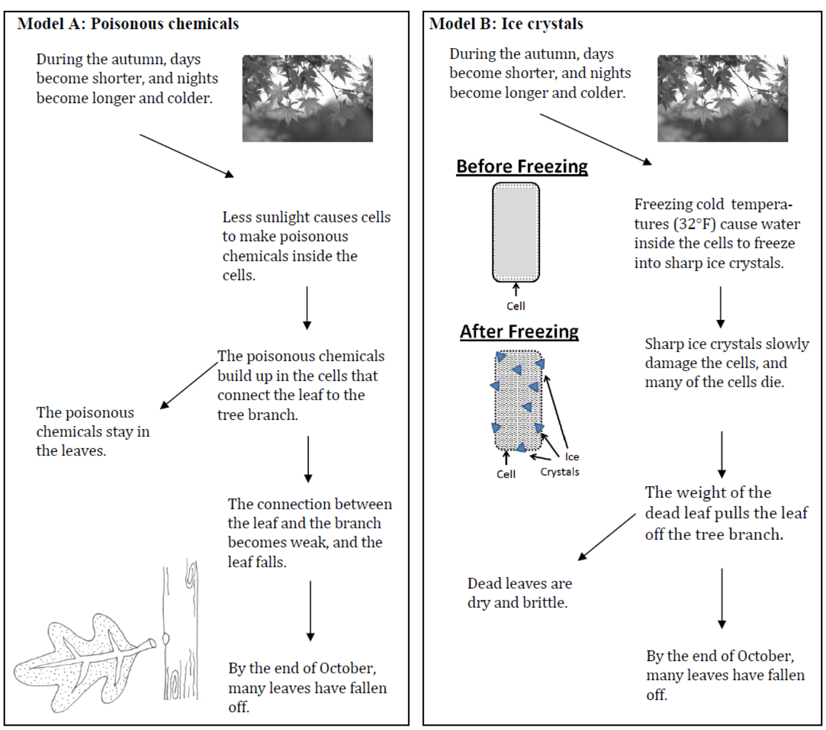
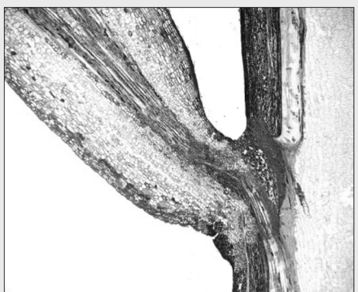


Figure 3.2. Competing models for falling leaves

Evidence #1: Leaves begin falling from trees in September. By late November, almost all of the leaves have fallen and the trees are bare.	**Evidence #2:** Leaves are made up of different types of cells. When leaf cells are damaged, the leaf may die.

Figure 3.3. Evidence 1-2 for falling leaves

<table>
<tr>
<td>Evidence #3: Scientists compared leaves from trees in the summer with leaves that had just fallen to the ground in the autumn. Nearly all the cells in green leaves in the summer were alive, but there were more dead cells in leaves that had just fallen.</td>
<td>Evidence #4: Scientists used a powerful microscope to take close-up photographs of the leaf stem. They took photos of the leaf stem during the summer, and then again during the autumn. Here are the photographs.

Leaf stem during the summer

Leaf stem during the autumn</td>
<td>Evidence #5: Scientists examined the percentage of leaves that fell off trees in New Jersey during an autumn with warm temperatures. This is what they found:
<table>
<tr><th>Month</th><th>Number of nights the temperature was below freezing</th><th>Total % of tree leaves that had fallen off by the end of the month</th></tr>
<tr><td>August</td><td>0</td><td>0%</td></tr>
<tr><td>September</td><td>0</td><td>14%</td></tr>
<tr><td>October</td><td>0</td><td>63%</td></tr>
<tr><td>November</td><td>2</td><td>87%</td></tr>
<tr><td>December</td><td>17</td><td>100%</td></tr>
</table>
The table shows that each month more and more leaves had fallen off the trees, until 87% of all leaves had fallen off by the end of November, and 100% had fallen off by the end of December.</td>
</tr>
</table>

Figure 3.4. Evidence 3-5 for falling leaves

Typical Lesson

The first lesson in the cells unit—What do chloroplasts do? – is an example of a typical lesson in the curriculum. In this lesson, students were provided with six pieces of evidence, most of these were on computers. Throughout the lesson, students used the evidence to create and revise models for the function of chloroplasts. At the start of the lesson, students read a paragraph on chloroplasts to provide some science content before they were introduced to the first four pieces of evidence. The science content provided at the beginning of the lesson did not discuss the function of chloroplasts. Each piece of evidence was placed in a specific order, and students were asked to complete various tasks per piece of evidence. The first four pieces of evidence provided information about chloroplasts to provide additional science content. After students constructed their initial models of chloroplast function and discussed them with their classmates, they read the final two pieces of evidence. Based on the new pieces of evidence, students revised their initial models.

The order and design of each piece of evidence were intended to support students understanding of the function of chloroplasts. Evidence 1, Elodea experiment, involved a hands-on experiment in which students observed chloroplasts in plant cells under a microscope at 40x and 100x and drew/wrote their observations. They individually answered an evidence understanding question, "what do you conclude from this experiment, remembering the observations made about the size, shape, colors, and patterns seen in the cells?" In Evidence 2, bacteria experiment, students read a research report on bacteria. The report stated that a scientist, Engelmann, wanted to see if oxygen was produced in a cell and where. He placed plant cells from a plant called Spirogyra on a microscope slide and then put bacteria that are attracted to oxygen on the slide and exposed it to light. Engelmann found that the bacteria went to the places around the edge of the cells next to the chloroplasts. The report did not include a conclusion for the experiment. Instead, students discussed the content of the report in pairs and individually wrote a conclusion for the report. Evidence 3, “Where chloroplasts are found?”, was a computer simulation that carried out by students. Students viewed slides of animal and plant cell tissues to see if they had chloroplasts. Students concluded that animal cells had no chloroplasts, cactus root and oak tree trunk cells had few chloroplasts and oak tree leaves, moss, tulips, cactus and algae cells had many chloroplasts. Evidence 4, the starch experiment, was another computer simulation. For this experiment, an indicator solution was applied to the leaves that turns the leaves blue in the presence of starch and brown in its absence. Students found that the green parts of the leaf turned dark blue and contained starch. The white parts of the leaf turned brown and contained no starch. The parts of the leaf covered-up turned brown and contained no starch. Students learned that the production of starch entails the production of glucose. At this point students, working in pairs, produced their initial models.

After students constructed their initial model, they read the final two pieces of evidence. In Evidence 5, “What makes chloroplasts work?”, students read a research report and in pairs discussed the report and individually wrote a conclusion for the report. Using a microscope, the authors of the report examined barley plants with normal chloroplasts and very mutated chloroplasts; and measured how much glucose is produced in each kind of barley plant. Students could conclude that barley plants with normal chloroplasts make large amounts of glucose; barley plants with mutated chloroplasts make small amounts of glucose; and barley plants with very mutated chloroplasts make no glucose. Evidence 6, chloroplasts and gases, was another computer simulation. The simulation demonstrated that chloroplasts absorb light and carbon dioxide to produce oxygen. The computer simulation allowed the pairs of students to determine whether chloroplasts use or produce carbon dioxide and oxygen. In the computer simulation, chloroplasts in a jar were placed in either dark or light conditions. The amount of carbon dioxide and oxygen in each jar was measured. For evidence 6, students found that in the light condition, the amount of oxygen in the jar got bigger, and the amount of carbon dioxide got smaller. In the dark condition, the amounts of oxygen and carbon dioxide did not change. After discussing the final two pieces of evidence, students revised their initial models. In groups, students explained their revised models to each other and discussed how well the models fit into the class's criteria for good models. Students wrote an explanation for why their model was a good model and what evidence supported their model.

As with the other units, lessons and assessments in this curriculum, the focus was on a driving question that students were asked to answer. Lessons introduced students to multiple pieces of evidence, which students used to create, evaluate, and revise models. The MEL matrix supported students as they evaluated models and evidence and helped them to decide how each piece of evidence supported, contradicted, or was irrelevant to the models the students evaluated. Students used the suite of scaffolds embedded in the curriculum to evaluate the models and evidence to provide an argument that supported one model and use evidence to support their reasoning. The particular lessons that focused on students' written arguments are discussed in more detail in the next section.

Focal Lessons

The focal lessons for this study were lessons in which students were asked to provide a written argument as one of their tasks for the lesson. These tasks were chosen for analysis because the writing prompts were similar to the pretest and posttest arguments students wrote before and after the intervention—that is, they focused on students evaluating models. For falling leaves the main question was, “which do you think is the better explanation of why leaves fall from trees to the ground during the autumn season—the Poisonous Chemicals explanation or the Ice Crystals explanation? Write **at least three** (3) detailed reasons for your answer.” The prompt, which was similar to the pretest and posttest question in the HIV lessons was, “write an argument to support the model you chose. Write to someone who may not agree with you.” Other variations of this writing prompt were used in the Genetics modeling assessment (“Which do you think is the best explanation of DEB, the separatin protein explanation or the connection protein explanation? Write **at least three (3)** detailed reasons for your answer.”), the Finches lesson (“Pick which model you think is better, Beak Size Model or the Fighting Model: Write an argument to support the model you chose. Write to someone who may not agree with you. Use your argumentation rubric to help you write your argument.”), and the posttest (“Which do you think is the better explanation of why leaves fall from trees to the ground during the autumn season—the Poisonous Chemicals model or the Ice Crystals model? Write **at least three** (3) detailed reasons for your answer.”).

The other type of question students answered came from the model construction tasks, but in their essays they evaluated the model they constructed. For instance, students were asked to, “individually explain why your model is a good model. Explain what evidence supports your model and explain in detail how the evidence supports your model.” Table 3.4 details the lessons in the curriculum that included written arguments analyzed in this study. During the curriculum, written arguments occurred during the

Cells, Genetics, and Evolution units--units 3, 4 and 5 respectively--and during specific lessons in each unit. For instance, in unit 4, Genetics - lesson 7: existence of HIV resistance, students provided a written response to a prompt that asked them to evaluate two competing models and use the body of evidence to support their model choice.

Table 3.4

Time Points for Students' Written Arguments

Assessments	Units	Lessons	Writing Prompt
Pretest (Muscle Pain/Falling Leaves)			Which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer.
	3. Cells	1. Chloroplasts: Students use various pieces of evidence in order to get an initial idea of the function of chloroplasts in plant cells.	Individually, explain why your model is a good model. Explain what evidence supports your model, and explain in detail how the evidence supports your model.
	3. Cells	2. Nucleus: Students determine the best model for the function of the nucleus, and revise the best model to make it even better based on what they learned from the pieces of evidence.	Individually, explain why your model is a good model. Explain that evidence supports the model, and explain in detail how the evidence supports your model.
	4. Genetics	7. Existence of HIV Resistance: Students learn to use scientific evidence to investigate the possibility that genetic resistance to HIV exists.	15. Write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer.
	4. Genetics	8. Mechanism of HIV Resistance: Students learn to use scientific evidence to investigate the specific cellular mechanism of genetically based resistance to HIV.	16. Write an argument to support the model you chose. Write to someone who may not agree with you.

Modelin g Assessm ent	4. Genetics		5. Which do you think is the best explanation of DEB, the separatin protein explanation or the connectin protein explanation? Write at least three (3) detailed reasons for your answer.
	5. Evolution	8. Finches: Students develop a model for the change in the finch population after the drought in 1977.	10. Pick which model you think is better, the Beak Size Model or the Fighting Model: Write an argument to support the model you chose. Write to someone who may not agree with you. Use your argumentation rubric to help you write your argument.
Posttest (Falling Leaves/ Muscle Pain)			Which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer.

Table 3.4 includes all writing prompts initially chosen for analysis, but not all classrooms completed all lessons. Each student, that was part of the sample, completed the pretest, posttest, nucleus lesson, genetics assessment and two additional writing prompts as part of other units. The additional lessons completed by students in this study varied by class.. Table 3.5 details the writing prompts completed for each pair. The pretest, posttest, nucleus lesson and genetics assessment were analyzed for all Students. The two additional writing responses were analyzed from lessons in the following units: Cells, Genetics, and Evolution. For pairs that were missing written responses to the Evolution unit, two writing prompts from the Genetics unit were analyzed. The reasoning for this was to pick a written response that were closest to the evolution unit to be able to identify changes that occur at different points and how aspects

of students' reasoning change during the course of the intervention. Appendix C provides additional details of the writing prompts, including a summary of the evidence presented for each prompt.

Table 3.5

Time Points for Written Arguments

Pairs	Pretest	Point 1	Point 2	Point 3	Point 4	Posttest
2 Pairs	Pretest	Chloroplast	Nucleus	HIV, lesson 7	Genetics Assessment	Posttest
8 Pairs	Pretest	Nucleus	HIV, lesson 8	Genetics Assessment	Evolution	Posttest

Below, I provide a brief summary of each focal lesson including the driving question for the lesson and a brief description of how the lesson was organized. The lessons are presented in the order they appear in the curriculum. See Appendix A and B for detailed explanations of each focal lesson including a lesson with modeling, content, and evidence understanding questions in Appendix B, specifically.

Unit 3 Cells, Lesson 1 Chloroplasts

This lesson was discussed in detail in the previous section. The driving question for this lesson was: "what do chloroplasts do?" Students learned about chloroplasts through six pieces of evidence. The purpose of the first four pieces of evidence was to provide students with background content about chloroplasts and students used the evidence in constructing their model. The last two pieces of evidence provided additional information about the function of chloroplasts and was introduced to students after they completed the tasks of revising their initial models and answering the writing prompt.

Unit 3: Cells, Lesson 2: Nucleus

This lesson was the second lesson in the cells unit, and the driving question was: “what does the nucleus do?” Students began this lesson by thinking about how “good” or “bad” the evidence was. Students were instructed that when evidence is “good”, you can believe the evidence conclusion, and when it is “bad”, you cannot. Students learned to rate each piece of evidence using the numbers 0, 1, and 2 for how good that piece of evidence was. By rating the goodness of evidence, students had a better understanding of differences in the quality of each piece of evidence and were able to identify the better quality evidence to support their model choice.

Students learned about two models, the control center model and the instructions model. Both models attempted to answer the question, “what does the nucleus do?”. Students were presented with the prompt, “here are two models of what a nucleus does. You will use three pieces of evidence determine which model is better. You will then revise the better model to make it even better. Read each one of the two models.” Students read five pieces of evidence, rated the quality of each piece of evidence individually (for evidence 1 and 2) and completed the arrows diagram (MEL matrix) for Evidences 1 through 5. For lessons that occurred after this lesson, the evidence quality rating was included in the MEL matrix and sometimes for individual pieces of evidence. Lesson 2 also emphasized discussing how a piece of evidence supported a model and explored the relationship between a model and evidence.

Evidence 1, Dolly the sheep, presented students with a computer simulation. In the simulation, students observed that when the egg cell nucleus of a black faced sheep was replaced with the fertilized egg nucleus of a white faced sheep, the resulting offspring had a white face. In Evidence 2, glowing cats, students read a Scientific American blog where scientists took the Green Fluorescence Protein (GFP) gene from a jellyfish and injected it into a cat's egg cell. The resulting kitten glowed under UV light just like jellyfish. With Evidence 3, clawed frogs, students read a report that some clawed frogs give birth to abnormal tadpoles that have many cells without a nucleus and found that these tadpoles die before becoming frogs. In Evidence 4, students read information about the function of the nucleus and found that the nucleus is the control center of the cell. Students read that the nucleus contains DNA that directs the functions of cell structures. Finally, with Evidence 5, diabetes, students read a study about diabetes and learned that eight out of 10 people with Type 1 diabetes have a mutation in their DRB gene resulting in no DRB protein production. Students further learned that nine out of 10 healthy people had no mutation in their DRB gene and thus produced DRB protein at normal levels.

After students read and evaluated each piece of evidence, they evaluated the two models. Students discussed in pairs which model is better based on the evidence presented and then individually revised the model chosen to improve the model's quality. Students explained their reasons for the changes they made and answered the question, "individually, explain why your model is a good model. Explain what evidence supports your model and explain in detail how the evidence supports your model." This was the same question as in the chloroplasts lesson, but in this lesson students evaluated two models, chose the better model, and revised the model.

Unit 4: Genetics, Lesson 7: Existence of HIV Resistance

The driving question for this lesson was, “does genetic resistance to HIV exist?” It began by introducing two models--genetics resistance to HIV does not exist and genetics resistance to HIV does exist--and asked students to select which model they thought was the better one. Students read four pieces of evidence, evaluated each piece of evidence, and completed the MEL matrix. Throughout this lesson, students answered comprehension questions to check their understanding of the evidence, rate the quality of the evidence, and after learning about Evidence 1 and 2, they evaluated the models and chose which model was better while providing reasoning for that choice. In Evidence 1, Feline Influenza Virus (FIV), students viewed a video in which a scientist found that many large wild cats had FIV in their blood but were not affected by it because of a genetic mutation that made them resistant to the disease. Whereas house cats do not have this genetic mutation and were susceptible to FIV. In Evidence 2, students read an article that concluded that humans were not resistant to HIV because HIV is transmissible in different ways. Based on Evidence 1 and 2, students may change their choice of model because they have new knowledge that provides reasoning that genetics resistance to HIV may exist.

With Evidence 3, Simian Immunodeficiency Virus (SIV), students read about four breeding experiments with eight parent monkeys. Scientists conducting the study found that if both parents were resistant to SIV, the offspring were resistant as well. If either or both parents were susceptible to SIV, the offspring were also susceptible. Finally, with Evidence 4, Dr. Paxton’s study, students read an article in which Dr. Paxton and a team of scientists studied a group of 25 people who had been exposed to HIV many times. Despite exposure the people in the study were HIV negative and had some resistance to HIV. After students read and discussed each piece of evidence, they filled out the MEL matrix and then completed a model evaluation task: “write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer.” Students used an argumentation rubric while answering the writing prompt.

Unit 4: Genetics, Lesson 8: Mechanism of HIV Resistance

The driving question for this lesson was, “why are some people resistant to HIV?” It began by introducing two models, attack and destroy model and keep it out model. Students read four pieces of evidence. As with lesson 7, the pieces of evidence ranged in quality and students evaluated each piece of evidence to complete the MEL matrix. In Evidence 1, the Burke family, students read a case in which everyone in the Burke family had been infected with HIV, except one child - Nikki. One member of the family believed that Nikki had a special gene leading to the creation of CCR5 protein, which kills HIV cells. With Evidence 2, comparison of white blood cells, students observed a computer simulation that compared the protein composition of white blood cells among people with and without HIV. The results of the simulation showed that people resistant to HIV did not have CCR5 protein on their white blood cells and those with HIV had the protein on their white blood cells. These two pieces of evidence contradicted each other, so when students evaluated the quality of the evidence they also had to think about the source of the evidence as well as the differences in the evidence to resolve the contradiction.

With Evidence 3, Hepatitis B and interferon, students read a health science news blog about how the body fights the hepatitis virus. The blog explained that Hepatitis B is a disease caused by a virus that attacks the liver. Interferon is a protein produced by cells when bacteria or viruses attack them and it activates immune cells to destroy viruses. The scientists in the blog treated 12 patients who were infected with Hepatitis B with Interferon and found after the treatment: three people had no Hepatitis B, five people had a very low number of Hepatitis B viruses in their blood, and four had a very high number of viruses in their blood. Finally in Evidence 4, Jen's favorite genes, students read a blog in which the blogger tested individuals resistant and susceptible to HIV for 4 different white blood cell proteins: CD3, CCR5, CCR7 and CD8. The blogger found that white blood cell membranes from people resistant to HIV did not have the CCR5 protein while membranes from those susceptible had CCR5. She also found that fluid leftover after taking out the membrane from both types of peoples' white blood cells had no CCR5 protein while the chemical mixture did. After students read and discussed each piece of evidence, they completed the MEL matrix and a model evaluation task: "write an argument to support the model you chose. Write to someone who may not agree with you." Students used an argumentation rubric while answering this prompt.

Unit 5: Evolution, Lesson 9: Finches

The driving question for this lesson was, "what happened to the ground finches of the Galapagos Islands?" Researchers studied a population of ground finches that live on the Galapagos Islands. A severe drought lasting several years caused a very large drop in the finch population. Students were asked to use evidence to determine how the population changed after the drought, and why the change occurred. Students worked in pairs and evaluated Evidence 1-5 on a computer. The five pieces of evidence provided background information about the Galapagos Islands, ground finches, predators that eat finches, food sources such as seeds, and differences in beak size before and after the drought. The order in which students studied the evidence was up to the pairs. After studying the five pieces of evidence, students individually developed a model. In Evidence 1, What is their island like?, students learned about the conditions of the island, Daphne Major, where the finches lived. Evidence 2 provided general information about group finches, including that finches eat a variety of seeds, live around 2 years, laid on average 3 eggs per month, and some genetic traits of the finches include beak size, wingspan, foot length, body color and beak color. With Evidence 3, predators on the island before and during the drought, students learned that a scientist found the number of predators before and during the drought was similar in number. In Evidence 4, seeds on the island before and during the drought, students learned that plant scientists' found that during the drought there was an increase in seeds with hard shells and a decrease of seeds with soft shells. With Evidence 5, differences in finches before and after the drought, students learned that scientists found that the number of finches and their weight decreased after the drought, but beak size, wingspan and foot length stayed the same. After studying the five pieces of evidence, students individually developed an initial model that showed how the population of ground finches changed after the drought and why. Students used the general natural selection model to support their own model development. After developing the models, students explained their models to each other in small groups. As part of the discussion, students addressed if their model fit with the class's criteria for good models and if there were changes they could make to their model to fit the criteria even better.

After the discussion of their initial models, students studied Evidence 6-9 in pairs. Evidence 6, variation in finch beaks and how the beaks are used, explained that a professor found that finches with bigger, more pointy beaks could crack open both small, soft seeds and larger, harder seeds. In Evidence 7, observation of birds fighting, students learned that a couple visited the island in 1978 and saw two birds fighting over a seed. The larger bird drove the smaller bird away and ate the seed. With Evidence 8, seeds from four plants before and during the drought, students learned that a research team found that during the drought the number and type of seeds diminished for medium and small, soft seeds, and the number of large, hard seeds was about the same before and during the drought. In Evidence 9, weight of finches that survived and died during the first 6 months of the drought, students learned that a research team found that birds that survived the drought were slightly heavier.

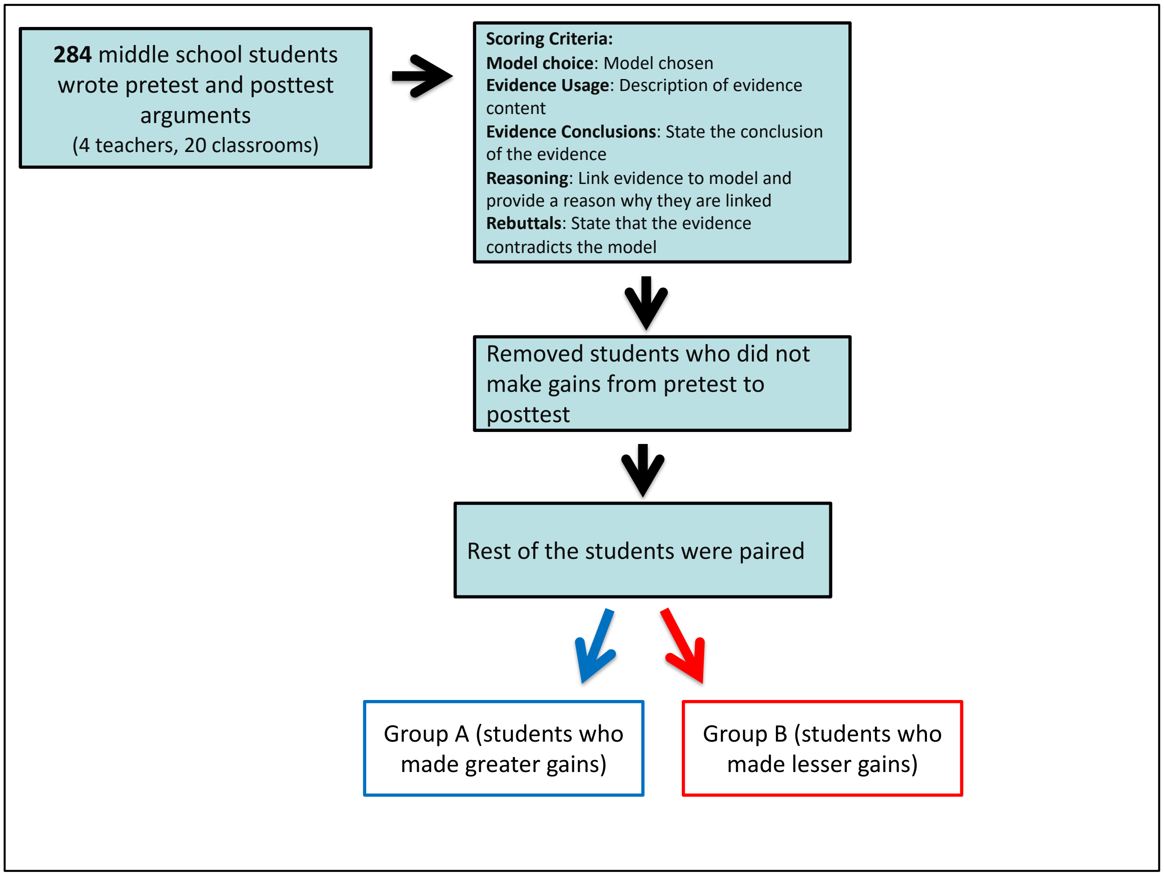
After students learned and discussed the final four pieces of evidence, the teacher introduced two models that could also explain what happened to the population of ground finches because of the drought. Students completed the MEL matrix and a model evaluation task. Students were prompted to, "pick which model you think is better, the Beak Size Model or the Fighting Model: Write an argument to support the model you chose. Write to someone who may not agree with you. Use your argumentation rubric to help you write your argument." This prompt was similar to the pretest and posttest writing prompt and provided students the opportunity to write an argument and use evidence to support their model choice.

By choosing similar writing prompts, it is possible to show how aspects of students reasoning may or may not have improved throughout the curriculum. It is important to choose students that were taught by the same teacher and in the same classroom with the only difference being that one student made greater gains at the end of the curriculum versus a student who made lesser gains. In the next section, I explain how the 20 students were chosen.

Participant Selection

The purpose of this study was to identify and track changes in students' written arguments before, during and after an intervention. Two groups of students, ten students each, were identified and compared from a sample of 284 seventh-grade students. The twenty students were purposefully selected using a scoring criteria that identified students who made gains from pretest to posttest modeling and argumentation assessments (e.g., muscle pain and falling leaves) (Patton, 2002). The two groups of students were paired into ten contrastive pairs with each pair having one student who made greater gains versus the other student who made lesser gains. Using a contrastive pairs approach to selecting participants is useful in this study because each contrastive pair included students who were in the same class and had the same teacher so the only difference for each pair is in the gains made from pretest to posttest. Figure 3.5 explains how the pairs were chosen.

As part of the pretest and posttest, students were introduced to two competing models and five pieces of evidence. The models attempt to answer the driving question of the assessment (e.g., Why do leaves fall?). One model was supported by more evidence than the other model. After being introduced to the driving question, students were asked to develop their own model, evaluate the two competing models and evaluate five pieces of evidence. Students answered the prompt: “which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer.” Students’ responses were coded, and the scoring criteria are explained in Table 3.6.



Paired students:

1. Students who made greater gains versus students who made lesser gains from pretest to posttest
2. Each pair had to have at least a five point difference
3. Each pair of students had to have the same teacher, be in the same class, and have written responses at the same points of the curriculum

Figure 3.5. How the pairs of student participants were chosen

Scoring each Pretest and Posttest

Table 3.6 explains the codes used to analyze students written responses. After analyzing students written responses, the responses were scored up to a maximum of eleven points. For a student to receive a maximum of eleven points, the student would receive one point for choosing the correct model, three points for either describing multiple pieces of evidence or providing an evidence conclusion (“other conclusion”, “other conclusion error” or “correct conclusion”), four points for reasoning about two or more pieces of evidence, two points for providing at least one rebuttal in the argument and one point for providing a correct evidence conclusion. The following is an example of a maximum score response provided by a student at the posttest:

> The Muscle Fibers model is the better model. Evidence 5 clearly states that the fibers aren't straight after exercise. This is due to the damage being done to the fibers that’s mentioned in Model B. Evidence 2 supports Model B because it says that the nerve cells attached to the muscles can sense changes. In Model B it says there is a new chemical and the nerve cells are irritated by it. Also Evidence 4 shows that people before and after exercise have an average of 1.5 lactic acid in cells. This proves Model A is wrong because it said lactic acid builds up and the cells swell up. But that is incorrect, the lactic acid is the same in the end after 24 hours. So as you can see Model B is the better model.

Table 3.6

Composite Argument Score Codes

Composite Argument Score Codes					
Type of Code	Pri-mary Code	Code Name	Definition	Example	Number of Points Assigned
Model Choice	A/B	Model	Student had chosen a model.	“I think that it’s the ice crystal that causes leaves to fall from the trees…”	0-1 Student received 1 point for picking the better model.
Evidence Usage	D	Describe	The student gave a detailed description of	"One reason why I think this is because	0-3

			the evidence content, with details about what was done, methods (if applicable), etc.	in evidence 5 it showed how when it was warmer out and there was more sunlight less percentage or leaves fell to the ground"	Student received points for either evidence usage or conclusion (OC, C, OCE)
Evidence Conclusions	OC	Other Conclusion	The student stated a conclusion with some basis in the evidence but failed to recognize the main point of the evidence.	"When you exercise you lose oxygen. The less oxygen the more lactic acid in the cells. When cells swell up from the lactic acid they push against nerve cells."	0-3 Student received points for either evidence usage or conclusion (OC, C, OCE)
Evidence Conclusions	OCE	Other Conclusion Error	The student stated a conclusion with some basis in the evidence, but there was some type of error in his/her understanding of the conclusion. The student also failed to recognize the main point of the evidence.	"I also think Model B is better because in evidence # 5 it showed on the table on the nights that the temperature was below freezing more leafs have fall on those nights then the other days"	0-3 Student received points for either evidence usage or conclusion (OC, C, OCE)
Evidence Conclusions	C	Conclusion	The student stated a conclusion that met the conclusion criteria. (For evidence 5: student mentioned that leaves fell off the tree even when the temperatures were not below freezing.)	"Thus there had to be another way: the poisonous chemicals. Explanation 2 is contradicted by evidence 5 in a way because in explanation 2 it said freezing cold temperatures cause it to fall of which didn't in fact happen until December after 95% of the leaves already fell off."	0-3 Student received points for either evidence usage or conclusion (OC, C, OCE) *If a student provided a correct conclusion the student received 1 additional point
Reasoning	R	Reasoning	The student linked a piece of evidence	"I think the Lactic Acid Model is a better	0, 2, or 4 points

			with the model and provided a reason why they were linked or what was linked in both model and evidence; explaining why the model supported the evidence.	model of why people get muscle pain 24 to 72 hours after heavy exercise. I think this because first, it shows how the muscles get swollen and Evidence #3 says the people who felt pain had swollen muscles."	Student received 2 points for reasoning about one piece of evidence, and 4 points if student reasoned about two or more pieces of evidence
Rebuttals	Ct	Contra-dicts	The student said or implied that the evidence contradicted a model.	"evidence 5 both supports model a and contradicts model b to prove that model a is the better model"	0 or 2 points Student received 2 points if the student used 1 or 2 pieces of evidence to contradict the model

Model Choice (0 - 1 points)

The first component for scoring the written responses was model choice. Model choice referred to which model was chosen by the student. A student who chose the better model based on the body of evidence, received 1 point. This code was included in scoring to capture instances when students understood the models, and were more likely to select the better model. I did not weight this score more than a single point because I did not want to overly penalize students who could provide good reasoning even though they picked the other model. The quality of responses depended on how the students interpreted the models, the body of the evidence, and relationship between model and evidence.

Evidence Usage and Evidence Conclusions (0 - 3 points)

The second component for scoring focused on how students used and/or interpreted each piece of evidence in their written argument. If a student provided a detailed description of one piece of evidence or understood the main points of the evidence, the student received one point. A student received up to three points for either providing a detailed description of three or more pieces of evidence or explaining the main point of three or more pieces of evidence.

Evidence usage identified if a student described a piece of evidence (e.g., the methods and any numerical results), whereas evidence conclusions identified if a student provided the conclusion for a piece of evidence and discussed it in their argument. Students were given credit for including any type of evidence conclusion, including correct conclusion (C), other conclusion (OC), and other conclusion error (OCE), because often students understood some part of the evidence and explained it in their argument. An argument was coded as correct conclusion (C) when the argument included the main points of the evidence. Other students might recognize a part of the evidence, but not the main points. If so, they were providing an “other conclusion” (OC) or “other conclusion error” (OCE). For instance, in the falling leaves task the correct evidence conclusion for Evidence 5 is that leaves fall off of trees even when the temperatures are not below freezing. If a student explained the main points of the evidence, it was coded as “correct conclusion” (C). If the student only stated that as the temperature decreases, the amount of leaves that fell increased, it was coded as “other conclusion” (OC). The student explained a part of the evidence but did not include the main points of the evidence - that most of the leaves fall before the temperature was below freezing. To be coded as “other conclusion error” (OCE), the student stated that leaves did not fall even when the temperature was below freezing or erroneously interpreted the table that was included with this piece of evidence.

The codes for evidence usage and evidence conclusion were combined into one category for scoring because both types of codes focused on students' understanding of the main points of the evidence. For each essay students received up to three points for this category. If a student did not describe any piece of evidence or any conclusion, the student received a score of 0 for this category. To receive a score of 1, a student needed to describe one piece of evidence or provide one conclusion. To receive a score of 2, a student needed to describe two pieces of evidence, provide two conclusions or describe one piece of evidence and provide one conclusion for another piece of evidence. For a student to receive the maximum three points, the student needed to describe three or more pieces of evidence or provide three or more conclusions.

Reasoning (0, 2 or 4 points)

The third scoring component was "reasoning". Reasoning referred to when students connected a piece of evidence to the model and explained why the model supported or contradicted the piece of evidence. It was complex a task and harder for students to do because it required students to use evidence to support their model choice. Students often struggle with providing reasoning for why they chose specific pieces of evidence (McNeill et al., 2006) or using sufficient evidence in their written arguments (Sandoval & Millwood, 2005). For this reason, students received two points if they provided reasoning once in their written response and four points if they provided two or more instances of reasoning in their written response. This component was weighted more heavily than the others due to its complexity and our understanding that it is harder for students to accomplish.

Rebuttals (0 or 2 points)

The fourth scoring component was rebuttals. Rebuttals provided evidence and reasoning to rule out a possible claim or counterclaim (Krajcik & McNeill, 2015). Rebuttals were identified when evidence contradicted the other model (not the model chosen) or that evidence contradicted the model due to evidence quality. Generating rebuttals was a complex task and harder for students to accomplish because they needed to use additional evidence and reasoning to make a claim for why a piece of evidence contradicted a model (Berland & McNeill, 2010). Due to the complexity of this component of the written argument, students received two points if they directly indicated or implied that any piece of evidence contradicted a model.

Correct Evidence Conclusion (0 - 1 point)

Finally, students received an additional point if they included at least one correct evidence conclusion. Because students already received a score for having evidence conclusions regardless of them being correct or with an error (C, OC, OCE), this score identified those students who understood the main point of the evidence correctly. Regardless of whether students included one correct conclusion for a piece of evidence or correct conclusions for multiple pieces of evidence, they received only one additional point.

After 284 pretest and posttests were coded, three independent coding pairs coded 20% of the data with at least 85% interrater reliability. I used the scoring criteria to provide a score for each pretest and posttest. Students could receive up to 11 points if they picked the better model (1 point), described three or more pieces of evidence or provided three or more evidence conclusions (3 points), reasoned about two or more pieces of evidence (4 points), used at least one piece of evidence to contradict a model (2 points), and provided one correct evidence conclusion (1 point). Students' pretest and posttest scores ranged from zero to 11 points and there was slight overall improvement in students' scores from pretest to posttest.

Pairing Students

The ten contrastive pairs of students were chosen from 284 students who completed both the pretest and posttest argumentation assessments and curriculum. To identify students who made greater gains versus students who made lesser gains, I examined their scores on both pretest and posttest and paired a student that made a greater gain from pretest to posttest with a student who made less gains from pretest to posttest. Table 3.7 shows the contrastive pairs of students. For instance, in pair 1 both students received 0 - 1 point on the pretest, but one student received 6 points on the posttest and the other student received 2 points - the two students were paired up. Each pair of students had the same teacher, were in the same class together, and had completed written responses at the same points along the curriculum.

Table 3.7

Pairs of Students

Pairs	Teacher	Period	Pretest Total Score	Posttest Total Score	Posttest-Pretest
1	IS	45	0	6	6
1	IS	45	1	2	1
2	IS	45	1	2	1
2	IS	45	4	11	7
3	JD	3	0	10	10
3	JD	3	1	5	4
4	JD	10	5	6	1
4	JD	10	3	9	6

5	JD	10	7	8	1
5	JD	10	0	8	8
6	JT	56	0	9	9
6	JT	56	1	4	3
7	JT	56	1	8	7
7	JT	56	1	3	2
8	JT	56	1	2	1
8	JT	56	0	6	6
9	PM	2	1	9	8
9	PM	2	5	6	1
10	PM	2	0	1	1
10	PM	2	2	9	7

Data Analysis

To analyze trajectories of change in students' written arguments and on how students' reasoning emerged, I analyzed the whole argument and then components of each argument (model choice, evidence usage, evidence conclusions, reasoning and rebuttals) to look for patterns in the trajectories. Tables 3.6 and 3.8 through 3.12 present the coding schemes used. Each table includes the code name, definition, and example of each code. Additional analysis completed after analyzing the written arguments focused on students' understanding of content, models, evidence, and evaluating the quality of evidence using the evidence rating scale (0 = bad evidence, 1 = satisfactory evidence, 2 = good or very good evidence). The additional analysis for model choice included how students discussed their chosen model (Table 3.9) and assessing model quality (Table 3.8). Table 3.10 explicates the codes used to capture students' reasoning in a more nuanced manner by analyzing the evidence-model connections in each written argument and if there was a relationship between components of the argument (Shea & Duncan, 2013). The additional coding schemes used in this study are described in detail in the following sections.

Data analysis began by comparing the pretest and posttest scores for the two groups of ten students -- the ten students who made greater gains in their written arguments (Group A) and the ten students who made lesser gains (Group B) -- to explain how the two groups differed. They were compared by plotting their mean scores for pretest and posttest. I then compared the students' written responses in between the pretest and posttest for both groups. I coded and scored the additional essays for the chloroplasts, nucleus, HIV lessons 7 and 8, and evolution lessons and genetics assessments using the same codes (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals) that were used to code the pretest and posttest. I then plotted the means per essay for each group using the same eleven-point scale. An independent coder coded 20% of the data with 90% interrater reliability. Patterns in argumentation were identified, as were general differences between Groups A and B in their written arguments across the curriculum. After comparing essays per group from pretest to posttest, I analyzed the components of the argument that examined specific argument aspects to identify patterns. I charted plot points for each group across all six arguments (pretest, four arguments during instruction, posttest). The final step using the initial argumentation codes was to chart plot points for each student across his/her six written arguments. First by using the eleven-point scale, and then additional charts using the same individual codes as used for group means.

After the initial coding and analysis between groups and for individual students, the next step was to focus on patterns that emerged to see if there were any additional codes that needed to be added for group and individual written arguments. This step included evaluating students' understanding of content, models, and evidence by analyzing students' responses to multiple-choice, true/false and short answer questions embedded in the lessons and posttest to determine if there were group differences in how many questions each group answered correctly. Additional codes that were added focused on argumentation quality included how students discussed the model chosen (Table 3.9), assessing model quality (Table 3.8), and evaluating the quality of the evidence. Table 3.10 includes a scale (levels 0 - 3) that evaluates the model-evidence connections in each argument.

To detect interesting differences, I looked for themes and patterns that emerged from the written responses coded and reported a detailed view of what was found after the written responses were coded. I looked for quotes from students' written work to support the themes that emerged (Creswell, 2007). After identifying themes, I created tables to sort the different codes and created counts to identify how many students follow a similar pattern across whole arguments and individual components of arguments. Patterns were established by looking at between group means for students' written responses across arguments and at individual students' written responses. To identify patterns, I examined similarities and differences in responses for students in each group. For instance, some patterns that arose within the data were evidence usage increasing across arguments, evidence usage fluctuating across arguments and evidence usage increasing and then plateauing. To identify a pattern, I looked at both groups and individual students to identify how many times students' written responses include particular codes and how the arguments changed at different points of the curriculum.

In addition, to identify nuance in students' reasoning across arguments, I read through each argument and assigned an evidence-model connection for each written arguments (Table 3.10). The coding and analysis of the written arguments helped provide a description of how students' written arguments changed which goes beyond what can be found if only analyzing pretests and posttests. This analysis explained what was needed during instruction to help support students during model-based inquiry activities and support their reasoning.

Content, Model, and Evidence Understanding

Throughout the curriculum, different lessons included questions that evaluated students' understanding of content, models, and evidence. Questions were short answer, multiple choice or true/false. Students' understanding of content, models, and evidence were analyzed by evaluating their responses to these questions—often it was if they provided a correct response. For instance, for evidence understanding there would sometimes be a question after students reviewed a piece of evidence. The evidence understanding question may ask students to write what they concluded about the evidence. If students understood the evidence, they would discuss the main points of the evidence. Appendix D includes the questions asked about content, models, and evidence in the lessons and posttest.

Content Understanding

Students' understanding of the content was evaluated with questions embedded into the nucleus lesson, the evolution lesson and in the posttest. The content understanding questions did not have anything to do with students' understanding of the model. For example, in the evolution lessons students were introduced to the driving question, the models and in pairs studied the first five pieces of evidence. After studying the evidence, the students were asked to check their understanding of the content by answering multiple choice questions about what they learned about finches (the topic of this evolution lesson) and the environment where the finches lived. After answering the multiple choice questions, students reviewed the next set of evidence. For the nucleus lesson, students' responded to a short answer question, "what is the function of the nucleus?" For the posttest, the content understanding questions were found in the middle of the test and included both short answer and multiple choice questions that focused on the chloroplasts and nucleus lessons.

Model Understanding

For analyzing model understanding, only one lesson (HIV lesson 8) included two multiple-choice model comprehension questions: (a) "According to the Attack-and-destroy model, why are some people resistant to HIV?", and (b) "According to the Keep-it-out model, why are some resistant to HIV?" Both questions were asked right after the two competing models were introduced to the students.

Another type of model understanding question included in the assessments and lessons asked students, "which model/explanation do you think is better? Circle one." In general, for lessons in the curriculum this question was asked at the end of the lesson right before students responded to the writing prompt, " write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer." For assessments, students were asked, "which model do you think is better? Circle your selection." Students were provided with three choices: Model 1, Model 2, or I can't tell right now. Whereas, for the lessons they were only provided with two choices: Model 1 or Model 2. To analyze these questions, I examined if students used the model chosen for this question when writing their final argument (the writing prompt) at the end of the lesson or assessment.

Evidence Understanding

To evaluate students' ability to understand the evidence, I evaluated evidence understanding questions that were embedded in the chloroplasts lesson, nucleus lesson, HIV lessons 7 and 8, and evolution lesson. The questions included multiple choice, true/false, and short answer questions. The questions were often asked right after students studied the evidence. Sometimes students were asked one question or more, it depended on the evidence. I analyzed the different evidence understanding questions to see if students: (a) understood the evidence (e.g., Did they answer the question correctly? or What new information did the student learn about the evidence?), (b) discussed the relationship between the evidence and model, and (c) how they responded to the question, "Which evidence is most useful for choosing a model?" For most of the lessons, students were asked to respond to questions in pairs, groups or individually. I focused on questions that asked students to respond individually so I could evaluate their understanding of the evidence.

Figure 3.6 illustrates an example of an evidence understanding question asked after students were introduced to the first piece of evidence in the chloroplasts lesson. This evidence understanding question asked students, "what do you conclude from this experiment?" The first piece of evidence was the results of the Elodea experiment. The evidence informed students that chloroplasts are green and found in plant cells.

Evidence #1 – Elodea Experiment

1. In pairs, complete Evidence #1. You will view elodea cells through a microscope and observe the chloroplasts inside the plant cells. The procedure for this experiment can be found next to the microscopes.

2. In the box provided below, individually write/draw your observations.

3. Individually, answer this question: What do you conclude from this experiment?

Figure 3.6. Example of evidence understanding question.

Coding Schemes

Reasons Citing Epistemic Criteria in Models (RECM)

Table 3.8 explicates the codes for reasons citing epistemic criteria in models (RECM). RECM falls under the epistemic list of criteria to assess epistemic products and practices, and includes both primary and secondary model quality codes. Primary epistemic criteria (RECM) are central to the epistemic practice of science. The primary RECM focus on communication and dissemination of model choice and include specifics as how accurate a model is, describing the model, or explaining why a model was chosen. The purpose of secondary epistemic criteria (RECM) focus on communication and dissemination of model choice and include specifics such as how the model was organized, if it had pictures, what details were included or if it made sense (Pluta et al., 2011). Students use these model quality criteria to evaluate the quality of the model and the criteria may help them in choosing the better model.

Table 3.8

Reasons Citing Epistemic Criteria in Models Codes

Reasons Citing Epistemic Criteria in Models			
Type of Code	Code Name	Definition	Example
OT	Other	Anything that did not fit below. Note: write down the "other" reasons in a separate list - high incidence reasons may be turned into codes.	
MS, MMS	Makes sense, Makes more sense	The model makes sense or makes more sense.	"The Lactic Acid Explanation makes more sense because the people without pain or swollen muscles rid their muscles of lactic acid."
CRI	Criteria	The student stated his/her model was good because it followed the criteria for good models.	"My model is a good model because it follows the criteria for good models; it is clear, broken down, organized, contains helpful pictures and relating language."
EX	Explains	The student stated his/her model was good because it explains.	"I think my model is good. I think this because it explains how the protein affects the physical appearance."
DE	Detail	The student stated his/her model was good because it included more detail.	"I just think it needs more detail and needs to mention what the nucleus produces."
PIC	Picture/Diagram/Charts	The student stated his/her model was good based on pictures or the quality of the those. Student used the term/phrase: picture, diagram, chart, visuals, drawing, image	"I think my model is good because it contains simple pictures that explain my model."

C	Clarity	The student stated his/her model was good because it was clear.	"My model is a good model because it is clear, short, simple…"
UN	Understand	The student stated his/her model was good because easy to understand.	"My model is good because it is easy to understand."
O	Organizatio n	The student stated that they were influenced in choosing a model based on whether/the degree to which it was neat and/or organized. Student used the term/phrase: neat, organized	"Another reason I think my model is good is because it's organized. All of my pictures are neat and you know where to look next."
EV	Evidence	The student stated that his/her model was good based on whether/the degree to which evidence was included and/or supported/contradicted the model. Student used the term/phrase: there is evidence (EV), all of the evidence supports the model (EVA), most of the evidence supports the model (EVM), none of the evidence supports the model (EVN), none of the evidence contradicts the model (EVC).	"I think the model is a good model because there are many evidences that support it."
RS	Reasoning	The student stated that his/her model was good based on whether/the degree to which it has a reasons/gives reasoning, or the quality of the reasons/reasoning. Student used the term/phrase: reasons, gives reasoning	"My model is a good model for many reasons. First of all, it has pictures to easily explain the model."
AC	Accuracy	The student stated that his/her model was a good model based on its degree of accuracy. Student used the term/phrase: accurate/inaccurate, correct/incorrect, true/false, right/wrong	"I think the beak size model is a better and more accurate model then the fighting model."

Explaining Why a Model was Chosen

Table 3.9 explicates the codes used to analyze why a student chose a specific model, students either: (a) used the model to explain why the model was chosen (MB),

(b) used the model to explain why the model was chosen, but did not understand the model (MBE), (c) chose a model due to evidence (ME), (d) chose a model due to higher quality evidence (evidence that supported the better supported model and/or contradicted the other model) (MEE), (e) chose a model because of model quality reasons (M-RECM), (f) chose a model because of model quality reasons and evidence (ME-RECM), or (g) did not discuss the model chosen (MN).

The first code was model choice. Model choices occurred when students used parts of the model to explain why they chose the model. The written argument focused on explaining the model and not using evidence to support their reasoning for choosing the model. The second code was model choice error. This code was similar to the model choice because students used parts of the model to explain why they chose the model but their explanation included model interpretation errors. With the next two codes, model evidence and model evidence (higher quality evidence), students used evidence to explain why they chose a model.

Table 3.9

Model Choice Codes

Model Choice Codes			
Type of Code	Code Name	Definition	Example
MB	Model Choice	Chose a model and used model to explain why model was chosen.	"I think the Ice Crystals explanation is better. Unlike the Poisonous Chemicals explanation, it provides a small diagram to help you to understand the explanation It shows how the ice crystals appear whe freezing."

MBE	Model Choice Error	Chose a model and used the model to explain why model was chosen, but does not understand the model. Error in understanding the model.	"The Lactic Acid Explanation makes more sense because the people without pain or swollen muscles rid their muscles of lactic acid. But if the damaged muscle fibers were the explanation, everybody who exercises would have damaged muscle fibers, and everybody would feel pain after exercising."
ME	Model Evidence	Chose a model due to evidence	"I think that the connectin protein is the better explanation because more of the evidence supports this explanation."
MEE	Model Evidence (higher quality evidence)	Chose a model due to higher quality evidence.	"Model 2 is the superior model. I think model 2 because evidence 2 and 4 supports it."
M RECM	Model Reasons Citing Epistemic Criteria in Models (RECM)	Chose a model due to reasons citing epistemic criteria in models (e.g., model is organized, clear)	"I think the Lactic Acid Model is the better Model. Model A is more detailed than Model B."
ME RECM	Model Evidence RECM	Chose a model due to evidence and reasons citing epistemic criteria in models.	"I think the beak size model is a better and more accurate model then the fighting model and I have many evidences to support it."
MN	No Model Discussed	Chose a model, but did not discuss why model was chosen.	"I believe that the Ice Crystals Model is the better model to explain why leaves fall during fall."

For model evidence, this included stating that most of the evidence supports the model to choosing a specific piece of evidence (e.g., Model A is the better model because evidence 2 supports it). The model evidence (higher quality) code (MEE) differs from the model evidence code (ME) because students used higher quality evidence to explain why they chose the model. Higher quality evidence is evidence that supports one model (the better supported model) and often contradicts the other model (lesser supported model). With the next two codes, model RECM and model evidence RECM, students used reasons citing epistemic criteria in models (RECM) to explain why they chose a model. For instance, for model RECM a student may write that the model is better because it is understandable, makes sense, is accurate, or is organized. Model evidence RECM differs from model RECM because students included evidence and RECM to explain why they chose the model. With the final category, no model discussed (MN), students did not explain why they chose the model. For this category, students would simply chose a model (e.g., Model B is the better model) but not use the model, evidence or qualities or criteria of the model to explain model choice.

Differences in Students' Reasoning

Table 3.10 explained the different levels of evidence-model connections in students' written arguments. Four levels of evidence-model connections were identified (levels 0 – 3). Level 0 was ascribed when no connection between evidence and model was identified in a student's argument. If an argument was identified as level 0 this meant that the student did not use any evidence to support the model chosen or refute the other model. Level 1 was ascribed when the written argument linked model and evidence, but did not provide a reason. For instance, "the model is supported by most of the models." Level 2 was ascribed when the written arguments included a reason, but the link between the evidence and model were superficial. The student did not justify how the evidence supported the model. The final level, level 3, was ascribed when the written argument included a reason and justification for why the evidence and model were linked. The final two levels, levels 2 and 3, were the only two levels that I identified as reasoning, but there were differences between these two levels. For a written argument to be scored at level 3, evidence-model connection, the argument needed to clearly state which evidence supported the model and why.

Table 3.10

Evidence-Model Connections

Evidence-Model Connections		
	Definition	Example
Level 0	No connection between evidence and model.	
Level 1	Provided an evidence-model connection, but no reason.	"I think model 1 because evidence 2 and 4 supports it."
Level 2	Provided a reason, but the link between evidence and model were superficial (did not justify why the evidence supports the model).	"Evidence 5 says that people with DEB are missing a protein which is what model 2 is saying." or "Model A and evidence 3 both say people who exercise hard will have swollen muscles."
Level 3	Provided a reason and justification for why the evidence and model were linked.	"Evidence 8 and 4 shows that after and during the drought, the amount of soft shelled seeds decreased greatly. This supports the beak size model because this is the reason that the model says the ground finches changed their beak size/smaller seeds." or "Evidence 5 also supports Model A. It supports it because the chart that it shows suggests that there is a different way than temperature freezing to make leaves fall."

Associations

My goal was to identify if there were relationships between the components of the arguments (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals). In addition, I wanted to determine if there was no relationship between two components, if one component was dependent on the other component or if components developed at the same time. Looking at the associations between components can explain if a student needs to understand one component before using the other component. Table 3.6 includes an explanation for each component and I focused on seven associations: (a) model choice and reasoning, (b) evidence usage and reasoning, (c) evidence conclusions and reasoning, (d) model choice and rebuttals, (e) evidence usage and rebuttals, (f) evidence conclusions and rebuttals, and (g) reasoning and rebuttals.

I began by identifying levels for each component and assigning points for each component for all 120 arguments. Tables 3.11-3.15 describe the levels for each component. Each component has a different number of levels that range from 0 - 2 to 0 - 4. Table 3.11 describes the levels for model choice for each argument. If the student chose the better supported model for the argument they would be at a level 1 and receive one point. If the student did not chose the better supported model they would be at a level 0 or received zero points.

Table 3.11

Levels for Model Choice

Levels	Model Choice
Level 0	Student picked the lesser supported model.
Level 1	Student picked the better supported model.

Table 3.12 identified the levels for evidence usage that range from 0 to 4. Evidence usage included: (a) mentioned evidence, (b) about evidence, and (c) described evidence. Mentioned evidence occurred when a student mentioned evidence by either number, name or picture (e.g., "in evidence #5 it says most of the leaves died when it was below freezing, which would have cause ice crystals"). About evidence occurred when a student gave a very general overview of the evidence content and stated what the evidence was

about (e.g., "The first reason the LAE says that the muscle feels pain because LA builds up in the muscles but after LA returned to normal the boy can definitely still feel pain"). Described evidence was explained in Table 3.6. Instances of described evidence occurred when, and it was when a student gave a detailed description of the evidence content, with details about what was done and, methods (if applicable). (e.g., "Dr. Stephen O'Brien noticed that house cats get FIV very quickly and die from it). An argument received zero points for evidence usage for a level 0, not discussing any evidence in the argument. An argument received one point for level 1, two points for level 2, three points for level 3, and four points for level 4. Table 3.12 describes the different levels for evidence usage and how they were assigned for each argument.

Table 3.12

Levels for Evidence Usage

Levels	Evidence Usage
Level 0	Student did not discuss any evidence in an argument.
Level 1	Student mentioned 1-2 pieces of evidence in an argument.
Level 2	(a) Student mentioned 3-4 pieces of evidence in an argument. (b) Student discussed about for 1-2 pieces of evidence in an argument. or (c) Student mentioned 2 pieces of evidence and 1 about in an argument.
Level 3	(a) Students discussed about for 3-4 pieces of evidence in an argument. (b) Students described 1-2 pieces of evidence in an argument. (c) Student mentioned 3-4 pieces of evidence and 1-2 abouts in an argument. or (d) Student mentioned 2-3 pieces of evidence, 1 about, and described 1 piece of evidence in an argument.
Level 4	(a) Students described 3-4 pieces of evidence in an argument. (b) Student discussed about for 1-2 pieces of evidence and described 2 pieces of evidence in an argument. or (c) Student discussed about for 3-4 pieces of evidence, and described 1 piece of evidence in an argument.

Table 3.13 describes the levels for evidence conclusions that range from 0 to 4. Table 3.6 defined the three different types of evidence conclusions: (a) correct conclusion, (b) other conclusion, and (c) other conclusion error. The levels for evidence conclusion correspond with the number of evidence conclusions provided in an argument. At level 0, the student did not discuss any evidence conclusions and receives zero points, and at level 4 for evidence conclusions an argument receives four points.

Table 3.13

Levels for Evidence Conclusions

Levels	Evidence Conclusions
Level 0	Student did not discuss any evidence conclusions in the argument.
Level 1	Student discussed 1 evidence conclusion in the argument.
Level 2	Student discussed 2 evidence conclusions in the argument.
Level 3	Student discussed 3 evidence conclusions in the argument.
Level 4	Students discussed 4 evidence conclusions in the argument.

Table 3.14 describes the levels for reasoning that range from 0 to 4. The levels of evidence-model connections (see table 3.10 for a description of each level) were used to identify the levels of reasoning. An argument receives zero points at level 0 if the student did not discuss any evidence-model connections in the argument, and at level 4 for reasoning an argument received four points.

Table 3.14

Levels for Reasoning

Levels	Reasoning
Level 0	Student did not discuss any evidence-model connections in the argument.
Level 1	Student included a level 1 evidence-model connection in the argument.
Level 2	(a) Student included a level 2 evidence-model connection in the argument., or (b) Student included 2 level 1 evidence-model connections.
Level 3	(a) Student included a level 3 evidence-model connection in the argument., or (b) Student included 2 level 2 evidence-model connections.
Level 4	(a) Student included 2 level 3 evidence-model connections in the argument., or (b) Student included 3 level 2 evidence-model connections.

Table 3.15 identified the levels for rebuttals that range from 0 to 3. Rebuttals included: (a) evidence that contradicted a model due to evidence quality or (b) evidence that supported one model and contradicted the other model. An argument received zero points or was at a level 0 when the argument did not include any contradictory evidence. An argument received three points for a level 3 argument.

Table 3.15

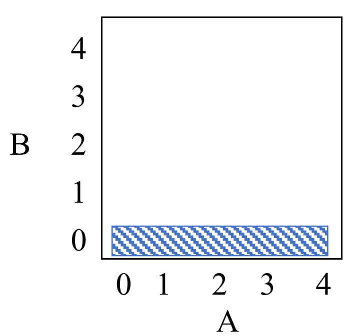
Levels for Rebuttals

Levels	Rebuttals
Level 0	Student did not include any contradiction in the argument.
Level 1	Student stated the evidence contradicted the model.
Level 2	Student stated the evidence contradicted the model and explained why the evidence contradicted the model.
Level 3	(a) Student stated the evidence supported one model and contradicted the other model. or

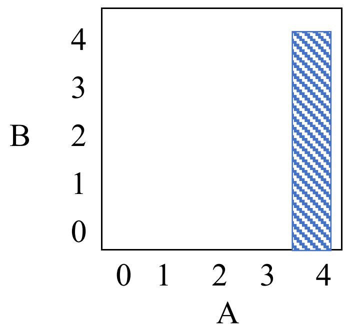
	(b) Student contradicted the evidence because the evidence was not good quality.

Figure 3.7 describes four examples of the associations that examined the relationships between two components of the argument: (a) A depended on B, (b) B depended on A, (c) parallel development, and (d) no relationship between A and B (Shea & Duncan, 2013). The point scale in the example figures were arbitrary for the associations analyzed in this study. The point scale was dependent on the scale for which the variable was measured. For each association between two components, the mean score of the levels or points were calculated for the 10 students in each group at each time point before, during, and after the curriculum (pretest, point 1, point 2, point 3, point 4, posttest). When the average score for each component was calculated, the two points were plotted in the graph for each group. Each graph had six points for arguments 1 - 6 and were labeled 1- 6 on the graph. After the points were plotted in the graph, I appraised the association between components.

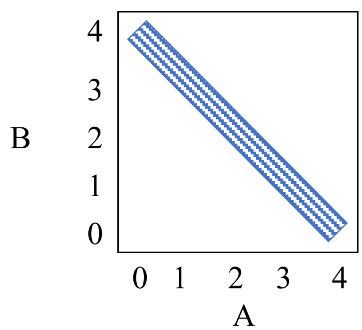
A Depended on B



B Depended on A



A and B Develop in Parallel



No Relationship Between A and B

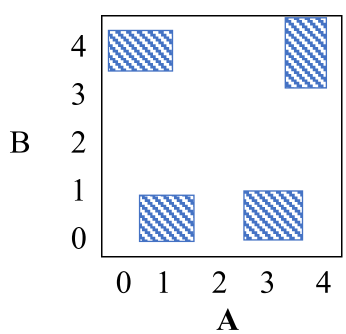


Figure 3.7. Four examples of associations

Inter-coder Reliability

First, all data was coded by an initial coder. Then, a second coder independently coded 20% of a random selection of data yet included arguments from each lesson and assessment. Across the coding schemes, the overall agreement levels between coders was 90%.

Significance

As part of a model-based inquiry curriculum, students should participate in the evaluation of competing models using multiple pieces of evidence. Also critical is students' understanding of how specific pieces of evidence support the model they choose. I wanted to understand and identify what strategies students use to evaluate models and evidence, and what additional supports are needed to help students as they reason about complex phenomena. This can include new approaches to instruction, including identifying supports and materials that help students become better at reasoning. Identified supports can be embedded into science curriculum. During implementation key questions arise such as, what aspects of model and evidence evaluation do teachers need to emphasize and make sure students understand?, and how do we assess students reasoning about key scientific content to support their evaluation and reasoning about key scientific content? Reviewing students' written work and examining how aspects of reasoning improved or declined throughout the intervention can help identify ways to support student learning and required teacher support. This study is especially timely and well positioned since authentic inquiry in science is complex and ways to support student learning in a model-based inquiry learning environment will hopefully lead to changes in how students evaluate data.

Chapter 4

Findings

This study investigated trajectories of change in students' arguments along multiple dimensions of reasoning before, during, and after a model-based inquiry curriculum. I focused on how students' reasoning emerged and changed over the course of a school year. As discussed earlier, I analyzed two groups: students who made greater gains from pretest to posttest (Group A) and students who made lesser gains from pretest to posttest (Group B). My purpose was to evaluate the similarities and differences in their written arguments. For both groups, I looked across arguments for trajectories of change in the structure and quality of their arguments. I also determined if they improved across multiple dimensions of reasoning. I looked at six arguments per student at different time points (before, during, and after the curriculum). Multiple time points were chosen for examination because evidence suggests that students' written arguments are likely to improve when they are given multiple opportunities to practice argument writing, evaluation of models and evidence, and additional time to discuss their reasoning.

The sample for this study (20 students) came from a larger sample (284 students). Initially, the larger sample's pretest and posttest were coded and analyzed, and I found students fell into 4 groups: (a) students who did not make any gains from pretest to posttest, (b) students who had the same score for pretest and posttest, (c) students who made lesser gains from pretest to posttest, and (d) students who made greater gains from pretest to posttest. For this study, I focused on the two latter groups in which students made gains from pretest to posttest. Although both groups made gains, their scores differed suggesting that similarities and differences between groups – in terms of argument structure and quality - may exist. This motivated the initial analyses of the arguments. For both groups, I analyzed: (a) how students' written arguments changed across arguments, (b) if certain competencies needed to be in place before other competencies emerged (e.g., do students include evidence in their arguments before providing a reason), (c) understanding of the content, models and evidence introduced for each lesson and the assessments, and (d) if supports introduced during the curriculum were helpful and led to changes in students' written arguments.

The 20 students included in the study were selected through a criteria explained in chapter 3. The students were split into two groups of 10 students each (Group A and B). For each student, I evaluated six arguments for a total of 120 arguments analyzed. Due to the small sample size, I did not run any tests for statistical significance. To examine similarities and differences between groups, I plotted the composite argument score and components of the argument for each student and group. I identified patterns in their arguments and explained how many students in each group fell into each pattern. The analyses focused on the research questions: (1) What are the trajectories of change over time in students' written arguments? (a) What are the differences in the quality and structure of students' arguments? And (2) In written argumentation, do certain competencies need to be in place before others develop? If yes, what are the competencies? For both questions I found that students' arguments most often changed over time, and there are certain competencies that needed to be in place before others develop, but it was group dependent.

To identify what is a good argument, I wanted to go beyond focusing only on the structure of an argument and look at other aspects to examine the quality of the argument.

I order to capture structure and quality features, I evaluated (a) students' understanding of models, evidence and content, (b) model and evidence quality, and (c) what criteria was used to assess the quality of the model. My purpose was to evaluate how higher quality arguments emerge, and explore if there are differences between groups in the development of higher quality arguments. To get a better understanding for the trajectories of change across arguments, I examined the two groups (A and B) and analyzed their (a) understanding of models, evidence and content to see if there were group differences in understanding, (b) different components of the arguments that included their model choice, evidence usage, evidence conclusions, reasoning and rebuttals, and (c) if parts of the arguments were associated with each other to identify if a certain competency needed to be in place before another competency emerged.

As part of the analysis, the groups' average scores for arguments and individual student's arguments were plotted for whole argument and for the components of the argument (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals) to look for trajectory patters. The patterns that emerged were different for trajectories in overall argument scores and in the different components of arguments - discussed in greater detail in this chapter. The last section of this chapter describes findings from the genetics assessment because there were differences in how students scored on this task versus other tasks that occurred before or after.

Identifying Patterns for the Composite Argument Score

In the first analysis, I examined the differences between the group that made greater gains from pretest to posttest (Group A), and the group that made lesser gains from pretest to posttest (Group B). For both groups, after all six essays were coded and analyzed per student; a total score out of 11 was given for each individual student's written argument at the different time points of the curriculum. The six essays included pretest, posttest, and four writing prompts across the curriculum.

The composite argument score included model choice, evidence usage and conclusions, reasoning, and rebuttals. Students could receive up to 11 points if they chose the better model (1 point), described three or more pieces of evidence or provided three or more evidence conclusions (3 points), reasoned about two or more pieces of evidence (4 points), used at least one piece of evidence to contradict a model (2 points), and provided one correct evidence conclusion (1 points). The average composite argument score for each group at each point was plotted to identify differences between the groups when focusing on how their overall scores across arguments. Figure 4.1 illustrates composite argument score at each point for the two groups.

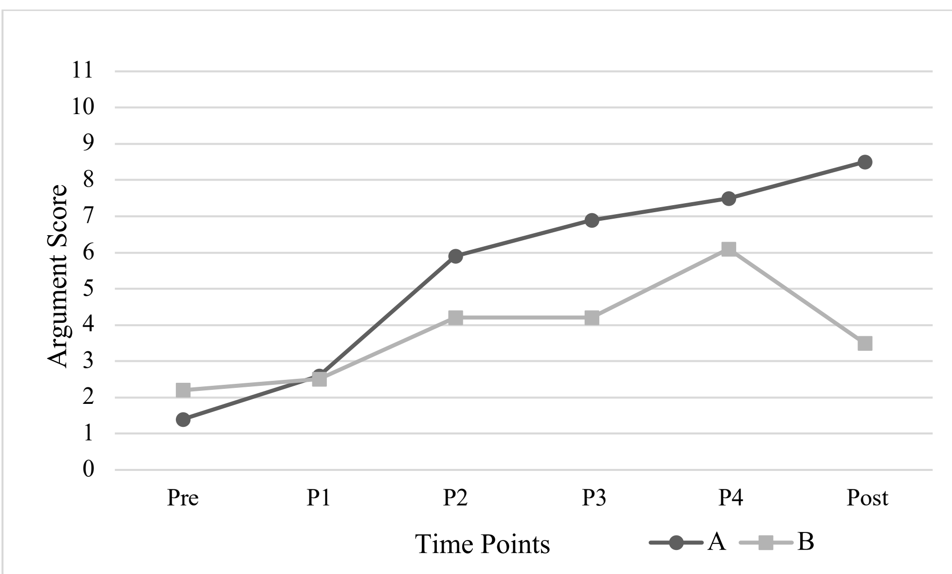


Figure 4.1. Composite argument score at each time point (P) for Groups A and B.

Although Group B started slightly better at pretest, Group A superseded Group B with Group A's written arguments continuing to increase across arguments. From pretest to posttest, Group A's total score continued to increase, but the trajectory for Group B was different. For Group B, there was a slower increase followed by a drop off at the posttest. At the posttest, Group B's written arguments differed from Group A's arguments because Group B provided less evidence conclusions, reasoning, and rebuttals in the final essay.

To examine if there were differences in both groups' written arguments by lesson or assessment, I plotted average total score for each group at each lesson and assessment. Figure 4.2 provides the plots for the two groups at each lesson and assessment. Across arguments, both groups started to increase by the middle of the curriculum (HIV lessons 7 and 8), but decreased at the genetics assessment. Group B had a slower increase with a drop off at the posttest. Group A continued to increase after the genetics assessment until the posttest.

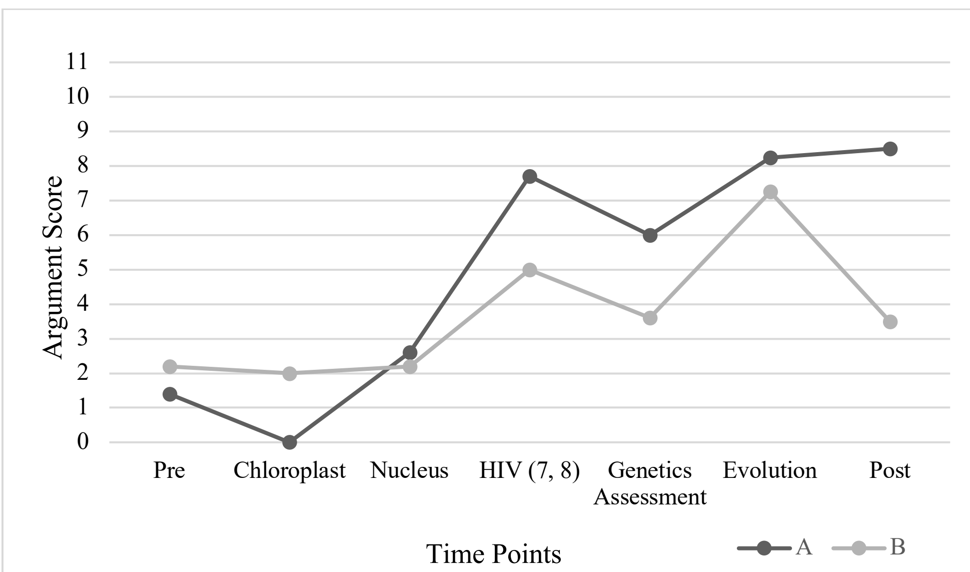
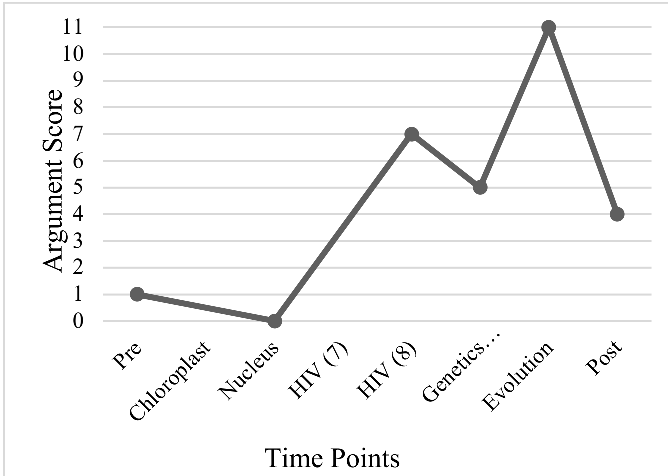


Figure 4.2. Composite argument score for each lesson or assessment.

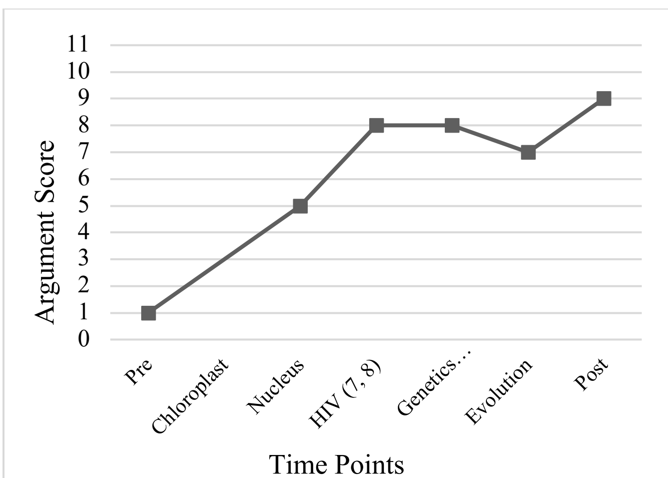
Although both groups showed some increase between pretest and posttest, the differences between the two groups was that Group A made greater gains (1.4 points at pretest to 8.5 points at posttest) than Group B (2.2 points at pretest to 3.5 points at posttest). Both groups' total scores dropped at the genetics assessment. Group A dropped from 7.7 points for the HIV lessons to 6.0 points at the genetics assessment. Group B dropped from 5.0 points for the HIV lessons to 3.6 points at the genetics assessment. Group B also had a decrease in total points between the evolution lesson (7.3 points) to posttest (3.5 points).

To determine what led to the differences between groups, the individual students' responses across arguments were plotted and some patterns for both groups emerged. Students in Group A and B had one similar pattern: arguments mostly increased but dropped at the assessments (genetics and/or posttest) (50% of Group A and 60% of Group B). Students in Group A fell into two other distinct patterns, (a) arguments increased monotonically from pretest to posttest (40% of Group A), and (b) arguments increased quickly and started to plateau on written arguments later in the curriculum and posttest (10% of Group A). Students in Group B fell into one other distinct pattern: argument quality stayed the same (40% of Group B). Figure 4.3a - d illustrates examples of students' arguments across tasks that follow the patterns described.

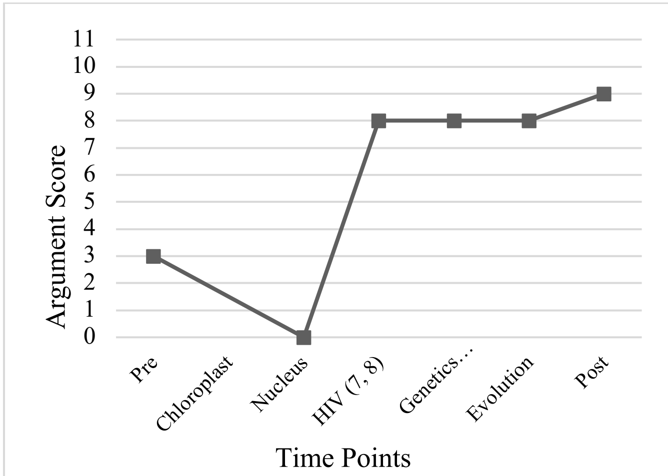
(a) Group A and B: Dropped at Assessment (Genetics and/or Posttest)



(b) Group A: Increased Monotonically



(c) Group A: Increased, Plateaued



(d) Group B: Stayed the Same Across Arguments

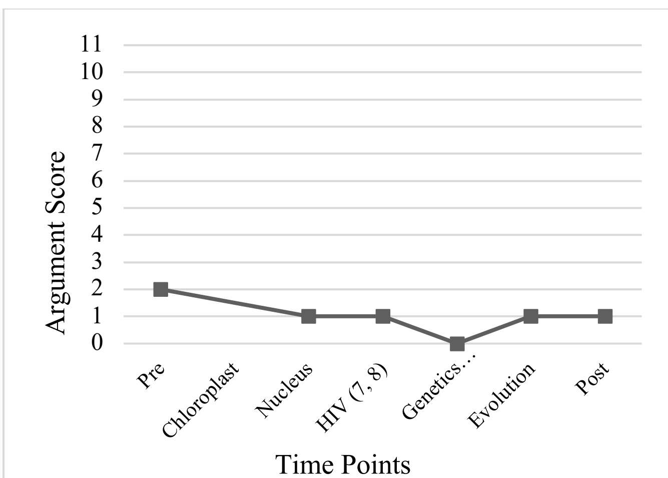


Figure 4.3a – d. Examples of the different types of arguments (composite argument score).

To provide a better understanding of why these patterns emerged, I analyzed both groups' understanding of models, evidence, and content. I then examined the components of the argument (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals). Next, I examined the components to identify if certain competencies needed to be in place before others. Finally, I examined why both groups' scores dropped at the genetics assessment.

Students' Understanding of Content, Models, and Evidence

Throughout the curriculum, different lessons included questions that evaluated students' understanding of models, evidence, and content. Evidence ratings asked students to evaluate the quality of the evidence. An explanation of the content, models, and evidence understanding questions are discussed in chapter 3 and Appendix D. Both the multiple choice, true/false and short answer questions for content, models, and evidence understanding were analyzed for accuracy. I began by analyzing the multiple choice and true/false questions. I did not find huge differences between groups in their understanding, so I then analyzed short answer responses to see if I found similar results for both groups. In this section, I begin with a discussion of students' understanding of content, models, and evidence to explain the similarities and differences between groups in students' understanding. In the final section, I discuss evidence evaluation focusing on how students in both groups rate the quality of evidence and how it relates to their usage of evidence. Evidence evaluation differs from evidence understanding because sometimes a student discussed the evidence quality but does not understand the evidence.

Content Understanding

Students' understanding of the content was evaluated with questions embedded into the nucleus lesson (three short-answer questions), the evolution lesson (six multiple choice questions), and the posttest (multiple choice and short answer questions from the chloroplast lesson and nucleus lesson). I found that students' understanding of the content slightly differed between groups with students in Group B having higher scores. This result was surprising because when evaluating students' whole arguments, students in Group A did better across arguments than students in Group B, which would suggest that students in Group A had a better understanding of the content. Because content understanding between groups was slightly different, I next discuss differences between groups in their understanding of models and evidence.

Model Understanding

Only one lesson (HIV lesson 8) included two multiple choice questions about model comprehension, for both groups, students were more likely to choose the correct responses to both questions. Another type of model understanding question included in the assessments and lessons asked students, "Which model/explanation do you think is better?" In general, students in both groups were more likely to discuss the model/explanation chosen in their written argument, but this differed for the genetics assessment and posttest. For the assessments, students were asked the same question mentioned above prior to being introduced to the higher quality evidence that either supported the better model and/or contradicted the lesser model. Students in Group A were more likely to choose the response "I can't tell right now" than students in Group B. For both groups, students who did chose this response were more likely to discuss the better supported model in their arguments for the genetics assessment and/or posttest. After seeing all the evidence, this helped students evaluate both models and led to choosing the better model.

Evidence Understanding

Evidence understanding questions were embedded in the chloroplasts lesson, nucleus lesson, HIV lessons 7 and 8, and the evolution lesson. The questions included multiple choice, true/false, and short answer questions. Figure 4.4 illustrates evidence understanding for the multiple choice and true/false questions across lessons.

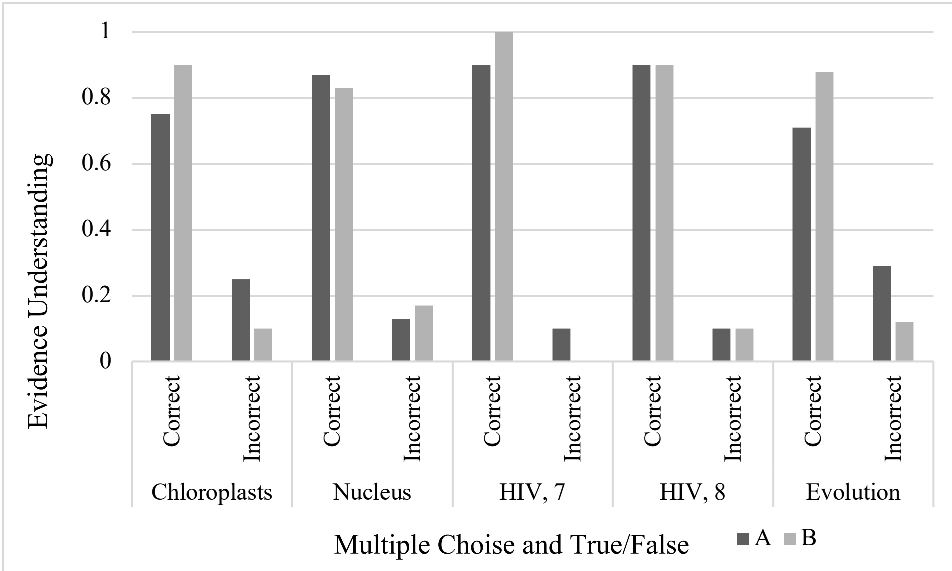


Figure 4.4. Evidence understanding (multiple choice and true/false).

When analyzing the multiple choice and true/false questions, students in Group B did slightly better in evidence understanding for the chloroplasts lesson and after the chloroplasts lesson, both groups had similar results. I then examined if there were similar results for evidence understanding for the short answer questions. Short answer questions were found in the chloroplast lesson, nucleus lesson, and HIV (lessons 7 and 8). I analyzed the different evidence understanding questions to see if students: (a) understood the evidence (e.g., did they answer the question correctly? or what new information did the student learn about the evidence?), (b) discussed the relationship between the evidence and model, and (c) how they responded to the question, “Which evidence is most useful for choosing a model?” (this question was only found for HIV, lesson 7). I found across lessons, students in both groups had similar responses except for the first evidence understanding question in the chloroplasts lesson. For the students that provided a response, there was a range of responses. Table 4.1 explicates the breakdown of student responses. Some students included more than one response in their answer. For instance, a student could include that chloroplasts are found in plants cells, are green, and discuss what chloroplasts look like.

Table 4.1

Evidence Understanding for Evidence 1 (Chloroplasts)

Group	A	B
Number of student responses	N = 8	N = 10
Chloroplasts are found in plant cells and are green	3	5
Discuss what the chloroplasts look like	4	8
Magnification leads to seeing more detail	4	5

For the rest of the lessons, both groups had similar responses to the questions. Most students understood the evidence and chose the higher quality piece of evidence to help them decide between models. For instance, for HIV lesson 7, most students in both groups picked evidence 4 (higher quality evidence). For their final arguments in this lesson, all four students who responded to this writing prompt discussed evidence 4 in their arguments. Three out of the four students (two from Group A and one from Group B) described evidence 4 and provided reasoning to support the better model.

The final short answer question asked students to evaluate the quality of a piece of evidence that was anecdotal and poor quality. Although the evidence was poor quality, most students in both groups (70% of Group A and 90% of Group B) rated the evidence as being good. Only two students in Group A thought the evidence was poor quality.

Although students mostly understood the evidence, they had difficulty evaluating the quality of the evidence.

Evidence Ratings

For both groups, I analyzed evidence ratings to examine how students' evaluate evidence and identify which evidence was higher in quality. I also determined students used the evaluated evidence in their arguments. In chapter 3, I explained the evidence ratings and evidence for each lesson and assessment. In both groups, students that rated the evidence as good or very good were more likely to discuss the evidence in their written arguments. There was a difference between evidence understanding and evidence evaluations earlier in the curriculum because a student can discuss evidence quality but did not understand the evidence. Earlier in the curriculum students that rated a piece of evidence as good or very good varied in their understanding of the evidence. Students in Group A understood parts of the evidence, which led to rating the evidence as good, but some had difficulty explaining the evidence conclusion. For Group B, most students rated Evidence 5, a piece of evidence that was higher in quality, as good or very good (90%) and had a better understanding of what the evidence concluded. Even though students in Group B had a better understanding of Evidence 5, more students in Group A (4 students) used Evidence 5 to support their reasoning for choosing a model than Group B (1 student).

By the middle of the curriculum, HIV lessons 7 and 8, both Groups A and B were similar in their understanding and rating of the evidence. However, how they rated the evidence did not always correspond with how they used the evidence. Evidence 4 was a higher quality pieces of evidence. For Evidence 4, students' rating of this evidence varied for both groups. Even though most students in Group A did not rate Evidence 4 as a good piece of evidence, most students (90%) either described the evidence, provided a correct conclusion and/or provided a reason in their written arguments. For Group B, four students that rated the evidence as good, either described the evidence, provided a correct conclusion and/or provided a reason in their written arguments. By the end of the curriculum, the evolution lesson, students' rating of evidence corresponded in their evidence usage. For both Groups A and B, students included evidence they rated as good in their written arguments. One difference between groups was that more students in Group A (75%) described more than one piece of evidence, provide evidence conclusions and provided more reasons than students in Group B (50%).

By the end of the curriculum, both Groups A and B had a similar understanding of models, content, and evidence. However, I found differences in students' overall argument quality between groups. Earlier in the curriculum, Group B had a slightly better understanding of content and evidence. In previous studies (e.g., McNeill, 2011; Osborne et al., 2004) argument quality was related to how well students understood the content. This suggests that Group B's arguments should be higher in quality. I did not find this. Even when Group had lower scores on content and evidence evaluation, the quality of their arguments were better than Group B. In the next section, I discuss the different components of the argument to examine other reasons for why students' arguments differed between Groups A and B.

Students' Usage of Models

For most of the lessons and the assessments, the final writing prompt asked students to pick the better model and use evidence to support the model chosen. To answer the writing prompt, students needed to evaluate the two competing models and multiple pieces of evidence that varied in quality. Overall, I found that students in Groups A and B understood the models and evidence, so differences in model understanding was not the reason why some students' chose the better supported model. I focused on how students used and discussed models in their written arguments to see if there were differences between groups in model choice and usage.

Model Choice

Model choice is defined as the model the student chose to discuss in his/her written argument. Figure 4.5 illustrates instances of choice of the better model across arguments for the two groups. Group A showed a greater increase from pretest to posttest in choosing the better model than Group B.

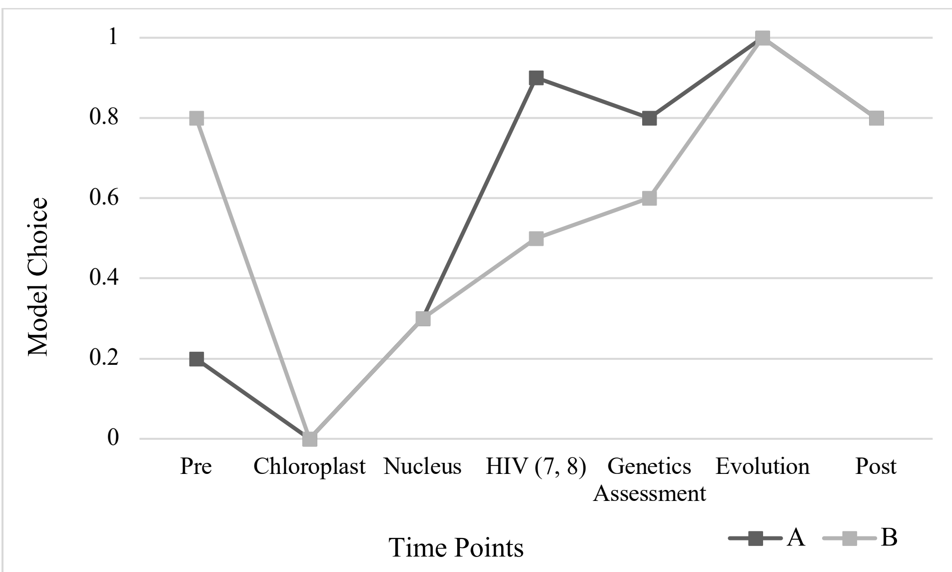
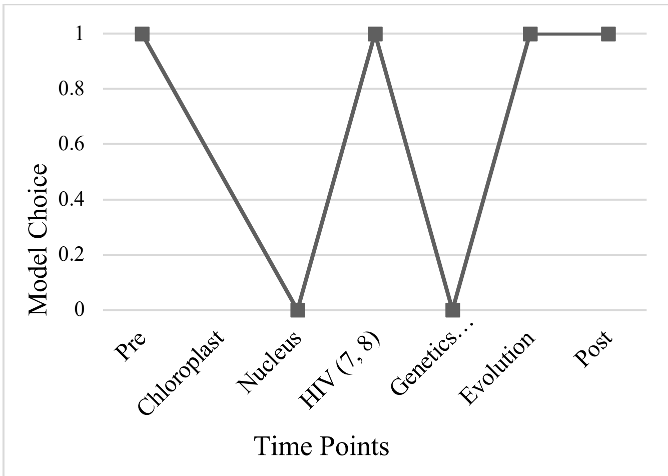


Figure 4.5. Model choice across arguments.

When looking at individual students' arguments across time, both groups exhibited similar patterns: (a) students chose the better supported model most often, but decreased at the genetics assessment or posttest (40% of students in Group A, 50% of students in Group B), and (b) students chose the better supported model most often (60% of students in Group A, 50% of students in Group B). Figure 4.6a and b illustrates examples of students' arguments across tasks that follow the patterns described.

(a) Dropped at the Genetics Assessments or Posttest



(b) Often Picked Better Supported Model

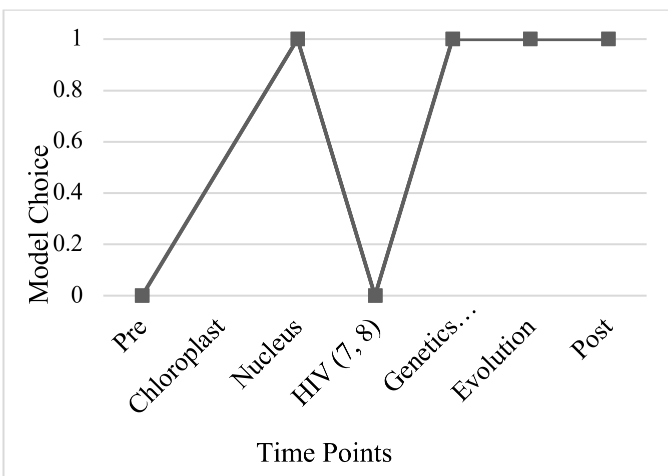


Figure 4.6a and b. Examples of the different types of arguments (model choice).

For model choice, students either were able to choose the better supported model most of the time or chose the less supported model at either the genetics assessment or posttest. This pattern was similar to the composite argument score where there was a decrease at the genetics assessment. A further examination of the genetics assessments is discussed later in this chapter. Because both groups have similar patterns, I analyzed the different ways students discussed the model chosen in their arguments to see if there were differences between groups.

Reasons Citing Epistemic Criteria in Models (RECMs)

Next I examined students' use of reasons citing epistemic criteria in models (RECMs). RECM falls under the epistemic list of criteria to assess epistemic products and practices, and includes both primary and secondary epistemic criteria. Primary epistemic criteria (RECMs) are central to the epistemic practice of science and focus on model accuracy to explain why a model was chosen or that the model explained, described or used evidence. Secondary epistemic criteria (RECMs) are more superficial reasons for model choice and included how the model was organized, if it had pictures, included details or made sense. Table 4.2 explicates the usage of epistemic criteria for each group.

Table 4.2

Reasons Citing Epistemic Criteria

		Pretest	Chloroplasts	Nucleus	HIV, 7 and 8	Genetics	Evolution	Posttest
Group A	Primary	0	100%	70%	50%	50%	20%	40%
	Secondary	10%	0	30%	0	20%	0	0
	Did Not Use RECM	90%	0	0	50%	30%	80%	60%
Group B		Pretest	Chloroplasts	Nucleus	HIV, 7	Genetics	Evolution	Posttest
	Primary	0	100%	70%	60%	50%	50%	40%

	Secondary	0	0	10%	0	0	0	30%
	Did Not Use RECM	100%	0	20%	40%	50%	50%	30%

In both Groups A and B, only half of the arguments included RECMs (52% of Group A and 55% of Group B). Arguments in each group that included RECMs were more likely to use primary RECMs than secondary RECMs (80% of Group A and 88% of Group B). When looking across arguments, epistemic criteria (RECM) were included more often and earlier in the curriculum. For the chloroplasts lesson, all students included primary RECMs. For the nucleus lesson all students in Group A (70% primary RECMs and 30% secondary RECMs) and most in Group B (70% primary RECMs and 10% secondary RECMs) used epistemic criteria to explain model choice. This made sense based on the writing prompts for these lessons. For instance, the nucleus lesson's writing prompt was, "Explain why your model is a good model. Explain what evidence supports the model, and explain in detail how the evidence supports your model."

Even though about half of the arguments included epistemic criteria to evaluate the model and explain reason for choosing the model, I felt it was important to analyze because it showed an aspect of how students evaluated models. Although both groups used similar amounts of RECMs (with Group B actually including more primary RECMs) and understood the models, there was a difference in their arguments based on evidence usage. Students in both groups understood most of the evidence, so evaluating how they used evidence helped to explain differences found between Groups A and B's whole arguments.

How Students Discussed the Model Chosen

In the methods section, I explained and identified six categories of students' discussions of the model chosen in their arguments. Table 4.3 explicates the breakdown of the different ways students discussed the model chosen. At the start of the curriculum, Group A's discussion of the model chosen began with using the model to discuss model choice (30% at pretest) or choosing the model due to evidence (40% at pretest, 20% at the nucleus lesson). By the midpoint and later in the curriculum, Group A's explanations shifted to choosing a model due to higher quality evidence (20% at nucleus lesson, 80% at HIV lessons 7 and 8, 50% at the genetics assessment, 62.5% at evolution lesson, and 40% at posttest). At start of the curriculum, Group B's discussion of the model chosen began with using the model to discuss model choice (70% at pretest) or choosing a model due to evidence (40% at pretest, 30% at nucleus lesson). At the midpoint and later in the curriculum, students in Group B continued to choose a model due to evidence (30% at HIV lessons 7 and 8, 30% at the genetics assessment, 12.5% at evolution lesson, and 20% at posttest) or choosing a model due to higher quality evidence (50% at HIV lessons 7 and 8, 40% at the genetics assessment, 62.5% at evolution lesson, and 30% at posttest).

Table 4.3

Discussion of Model Chosen

Lessons/Assessment		MB	MBE	ME	MEE	M RECM	ME RECM	MN
Pretest	A	30%	10%	40%				20%
	B	70%			30%			
Chloroplast	A							100%
	B							100%
Nucleus	A			20%	20%	20%	10%	30%

	B	10%		30%		30%		20%
HIV (7, 8)	A				80%		20%	
	B			30%	50%	20%		
Genetics	A			20%	50%	20%		10%
	B			30%	40%	20%	10%	
Evolution	A	12.5%			62.5%	12.5%	12.5%	
	B			12.5%	62.5%		12.5%	12.5%
Posttest	A			30%	40%	10%	10%	10%
	B			20%	30%	20%	10%	20%

Students in both groups understood the models and improved to using evidence to explain model choice over time. However, there was a slight difference between groups. Students in Group A were more likely to start using evidence earlier to explain model choice and moved to using higher quality evidence by the middle of the curriculum. It took students in Group B a longer time, towards the end of the curriculum and posttest, to use higher quality evidence to in their explanations. This suggests a difference in the groups' argument quality because part of evaluating argument quality is the student's ability to assess and evaluate the quality of both models. Model quality includes understanding the model, how students explain why they chose a particular model and the use of an epistemic list of criteria to evaluate the model (reasons citing epistemic criteria in models, RECMs). When students understood the models, chose the better-supported model and use criteria such as a "good model explains the evidence" a higher quality argument resulted versus choosing the lesser-supported model and using parts of the model to explain model choice. In the next section, I analyzed students' evidence usage.

How Were Students' Using Evidence?

The next component analyzed was how students used evidence in their written arguments. Each lesson and assessment included four to six pieces of evidence that varied in quality. The prompts directed students to evaluate and use evidence to discuss the model chosen in the argument. Although both groups had a similar understanding of the evidence, I evaluated students' evidence usage, evidence conclusion and what evidence students used in their arguments to identify any differences between groups.

Evidence Usage

Evidence usage occurs when a student gives a detailed description of the evidence content with details about what was done to collect or analyze the evidence (e.g., the methods and any numerical results). Figure 4.7 illustrates evidence usage for both groups and shows that students in Group A mostly increased but decreased at the assessments. Group B started off slightly higher than Group A, but across most of the arguments stayed about the same with only a slight increase at the evolution lesson.

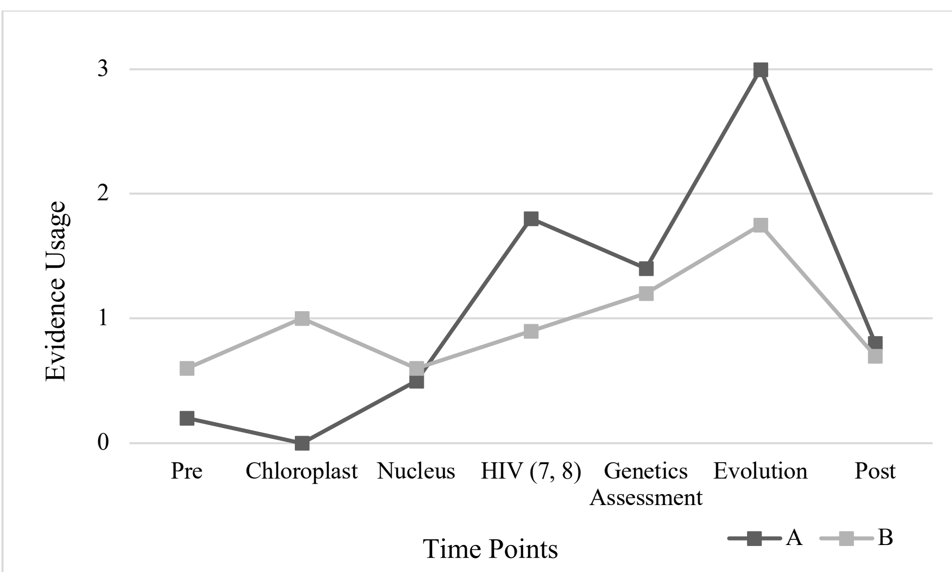
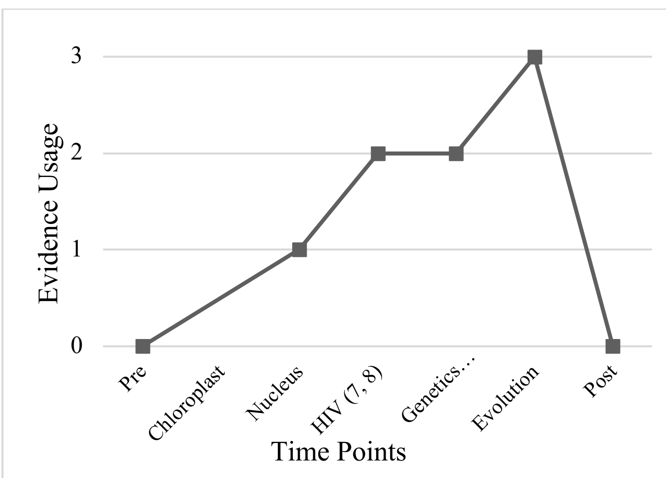


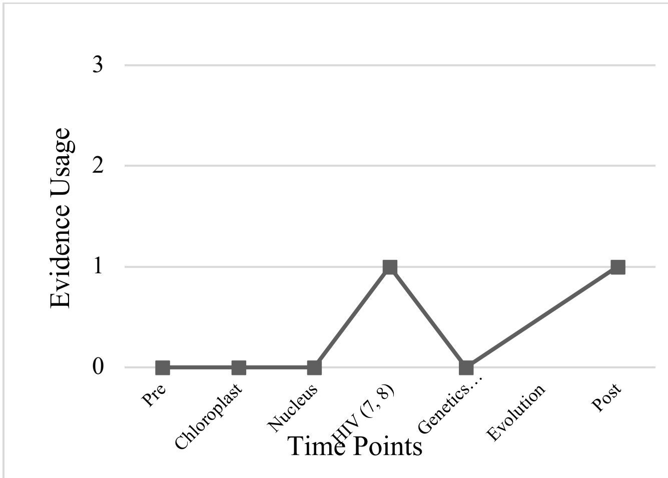
Figure 4.7. Evidence usage across arguments.

Figure 4.8a - e provides examples of the different evidence use patterns. Most students in Group A and B exhibited a similar pattern: students' evidence usage increased and then dropped at either the genetics assessment and/or posttest (80% of Group A and 60% of Group B). Two additional patterns were exhibited for Group A: (a) students only described a piece of evidence at one of the HIV lessons and then at the posttest (10% of Group A), and (b) student's evidence usage increased across arguments (10% of Group A). Two additional patterns were exhibited for Group B: (a) students did not describe any piece evidence across arguments (30% of Group B), and (b) student's usage of evidence slightly increased and then plateaued (10% of Group B).

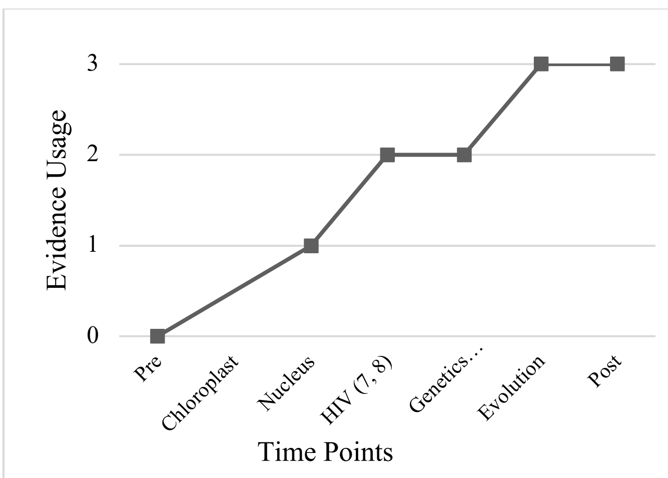
(a) Group A and B: Increased, Dropped at Genetics Assessment and/or Posttest



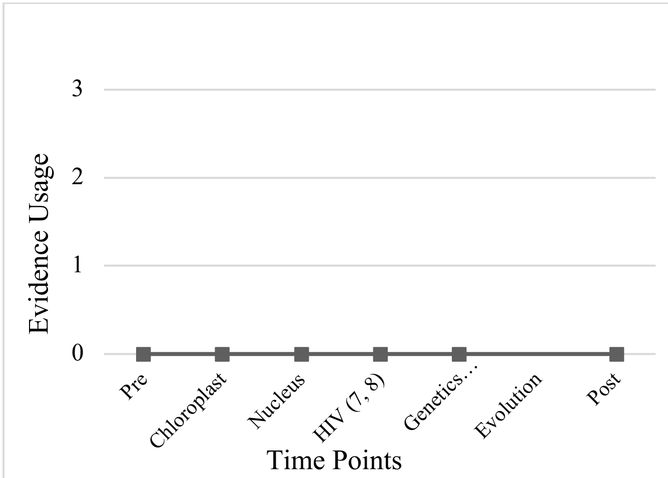
(b) Group A: Described Evidence at HIV and Posttest Arguments



(c) Group A: Increased Across



(d) Group B: Did Not Describe Evidence



(e) Group B: Slight Increased, Plateaued

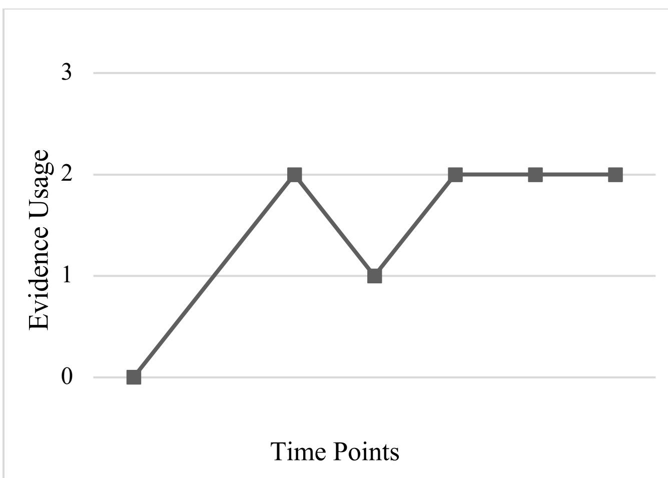


Figure 4.8a – e. Examples of the different types of arguments (evidence usage).

Evidence Conclusions

The analysis of evidence conclusions focused on the quality of students' statement of the conclusion of the evidence. As discussed in chapter 3, three types of evidence conclusions were identified: correct conclusion (C), other conclusion (OC) and other conclusion error (OCE). A correct conclusion occurred when a student stated a conclusion that met the conclusion criteria. This occurred when a student understood the main points of the evidence. Other conclusion occurred when a student stated a conclusion with some basis in the evidence but failed to recognize the main point of the evidence. Other conclusion error occurred when a student stated a conclusion with some basis in the evidence but there was some type of error in his/her understanding of the conclusion. The student also failed to recognize the main point of the evidence.

To provide better understanding of the differences in the two groups' usage of the three types of conclusions, all three types of conclusion were plotted onto one graph. Figure 4.9 illustrates all three types of conclusions for both groups. Across arguments, students in Group A increased in providing a correct conclusion and decreased in providing another conclusion or other conclusion error. For Group B, there was a slight increase in providing a correct conclusion across arguments but it was lower than Group A. By the evolution lesson, Group B decreased in providing other conclusion and other conclusion error and increased in providing a correct conclusion.

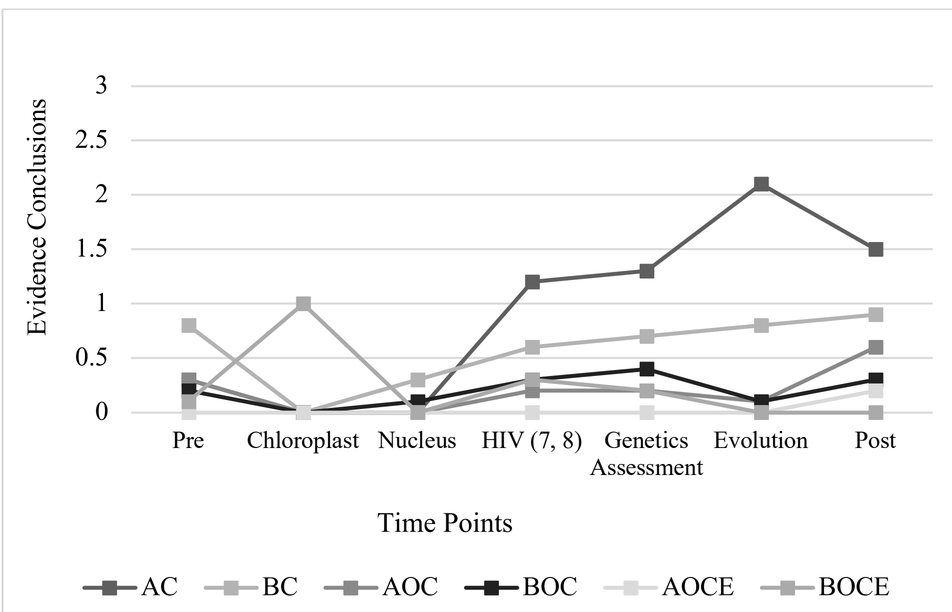
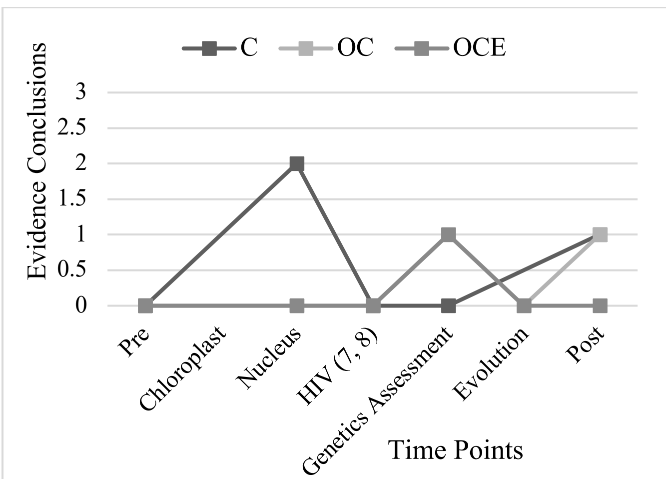


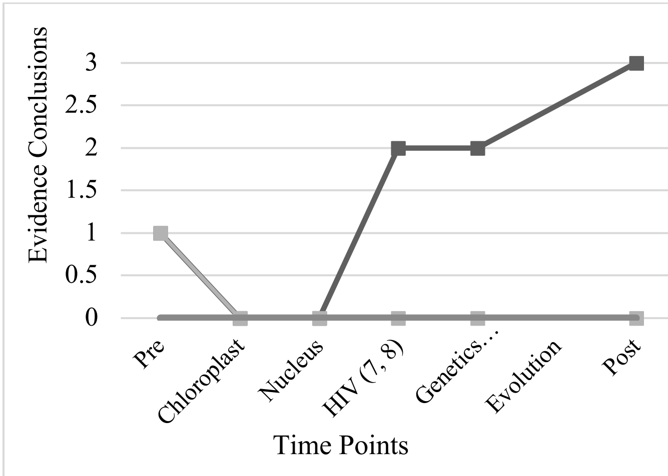
Figure 4.9. Evidence conclusions across arguments.

Figure 4.10a-d illustrates examples of the different patterns in evidence conclusions. Both groups exhibited two patterns: (a) students fluctuated in their usage of evidence conclusions (50% of students in Group A and 30% of students in Group B), and (b) students increased in using a correct conclusion, decreased in other conclusion and other conclusion error (50% of students in Group A and 30% of students in Group B). Two additional patterns were exhibited for Group B: (a) students only provided an other conclusion and/or other conclusion error (30% of Group B), and (b) students did not provide an evidence conclusion in any of the arguments (10% of Group B).

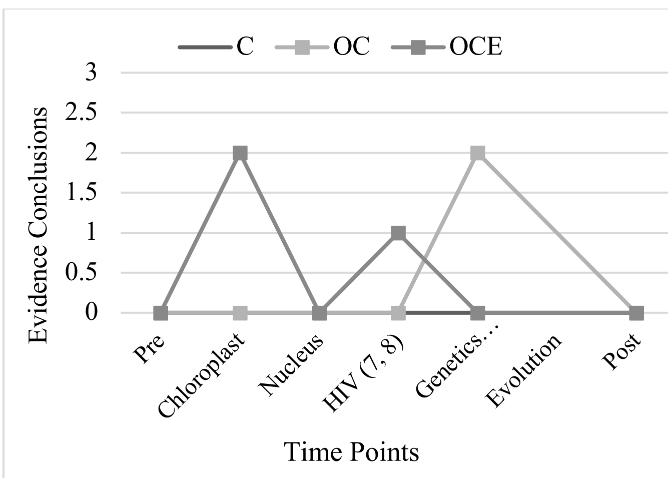
(a) Group A and B: Fluctuation in Usage



(b) Group A and B: Increased in Correct Conclusion



(c) Group B: Only Provided an OC or OCE



(d) Group B: No Conclusion

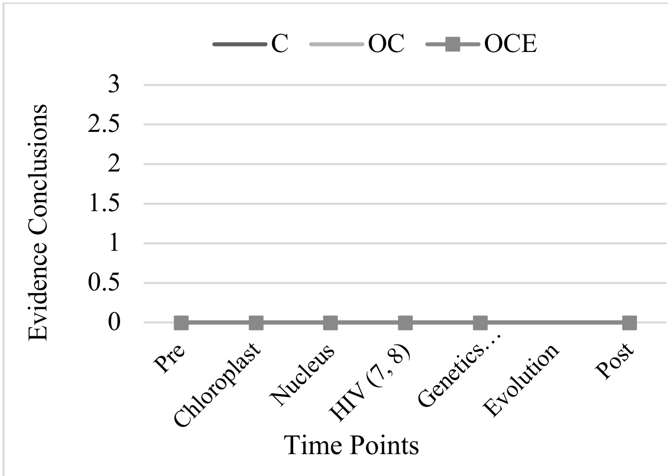


Figure 4.10a-d. Examples of the different types of arguments (evidence conclusions). For both groups, but especially in Group A, there was an increase in the usage of correct conclusion, and decrease in other conclusion and other conclusion error. This suggested that students' explanation of evidence improved across arguments and led to exploration of how students used evidence in their arguments.

Evaluating How Students Used Evidence

Evaluating how students used evidence is an important feature of written arguments because it is another component of argument quality. As an epistemic ideal, not all evidence is equal so that higher quality evidence is better (Chinn et al, 2014). Higher quality evidence is a reliable piece of evidence and includes facts, information from experts and often supports the better model and contradicts the lesser model. Poorer quality evidence is often anecdotal, does not include facts and supports the lesser model. Figure 4.11 illustrates higher quality evidence described in written arguments for each group.

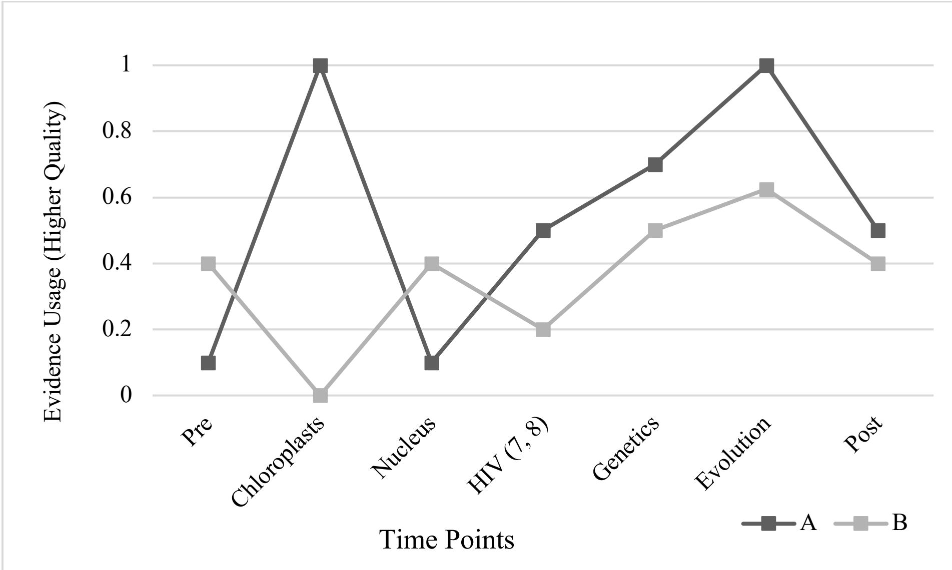


Figure 4.11. Higher quality evidence (described).

Across arguments, Group A included more mentions and detailed descriptions of higher quality evidence in their written arguments (73%) than Group B (57%). Group A's usage of higher quality evidence increased across arguments. By the middle of the curriculum most students described higher quality evidence in their arguments (50% for the HIV lessons and 70% the genetics assessment). By the end of the curriculum (the evolution lesson) every student described at least one piece of higher quality evidence in their written argument (100%). Across arguments, Group B's usage of higher quality evidence fluctuated. At pretest 40% of written arguments described higher quality evidence, by the middle curriculum 20% of Group B described the higher quality evidence in their arguments for the HIV lessons and 50% for the genetics assessments, and by the end of the curriculum (the evolution lesson) 63% of the students described one piece of higher quality evidence. For the posttest, I found that both groups decreased in their usage and description of higher quality evidence (50% of Group A and 40% of Group B).

Although students in both groups understanding of the evidence was similar, there were differences in how students used evidence. In general, more students in Group A (36%) described and used evidence to support the better model than students in Group B (29%). Across arguments, students in Group B used more evidence for the pretest than Group A, but evidence usage and how they used evidence differed between groups. Group B's evidence usage fluctuated across arguments, and Group A's evidence usage increased. Next, I analyzed students' reasoning across arguments to determine what role reasoning played in the development of argument quality.

Reasoning

My analysis for reasoning focused on students' linking a piece of evidence to a model or explaining why the model supports or contradicts the evidence. Figure 4.12 illustrates reasoning across arguments for both groups. Although Group A started out slightly higher than Group B at pretest, both groups follow a similar pattern for reasoning until the posttest, where Group B decreased and Group A increased. Group A made greater gains across arguments and students in this group were more likely to provide a reason during the HIV lessons, the evolution lesson, and this continued into the posttest. For Group B, students provided less reasons in their arguments before their use of reasoning dropped at posttest. I found that by the middle of the curriculum, Group A was more likely to use the higher quality evidence when providing a reason in their argument (80%) than Group B (40%). Towards the end of the curriculum and posttest, students in Group A provided more reasons and used higher quality evidence to support the model chosen (84% of written arguments) than Group B (35% of written arguments).

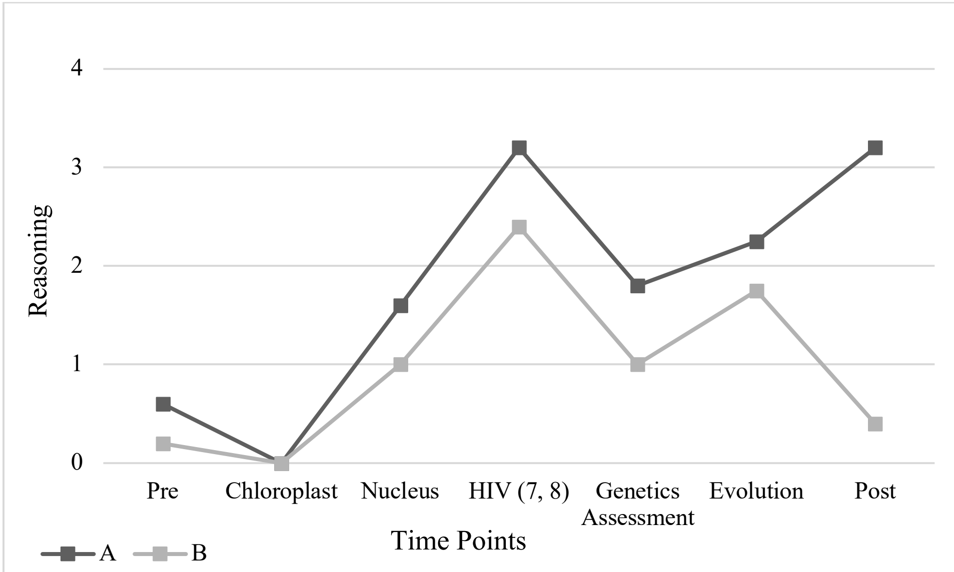
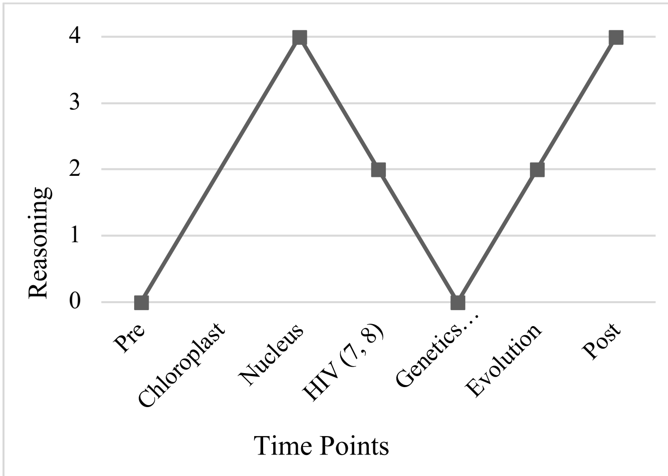


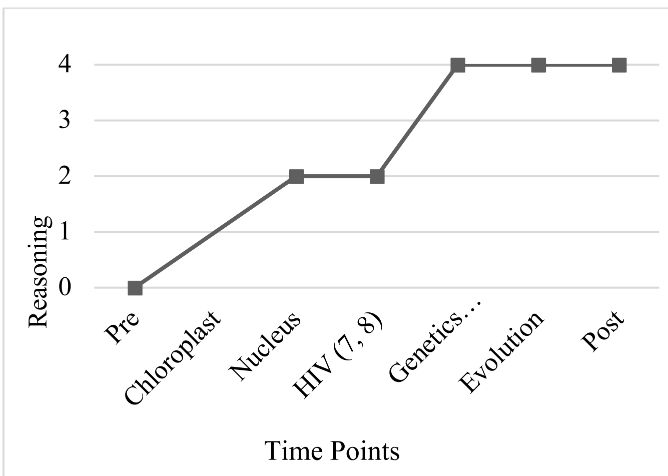
Figure 4.12. Reasoning across arguments.

When examining both groups, students' provided reasons followed two patterns: (a) students' provided reasons increased in complexity with a decline at either genetics assessment or posttest (50% of Group A and 30% of Group B) and (b) students' provided reasons increased with no gap (first argument with reasoning, the next arguments included reasoning and would increase towards the end of the curriculum) (50% of Group A and 20% of Group B). Students in Group B also exhibited one additional pattern: (a) students provided a reason for only one argument (30% of Group B) or did not provide any reasons (20% of Group B). Figure 4.13a-c illustrates an example of the three patterns.

(a) Group A and B: Reasoning Declined at Genetics Assessment or Posttest



(b) Group A and B: Reasoning Increased



(c) Group B: Provided Only One Reason

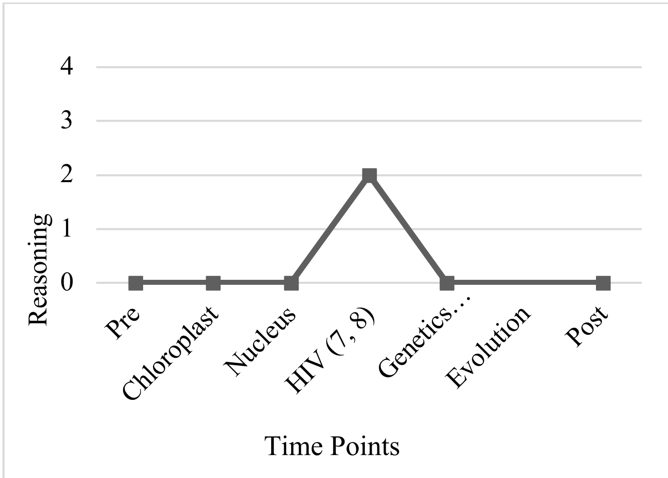


Figure 4.13a-c. Examples of the different types of arguments (reasoning).

Differences in Student's Reasoning

Reasoning mostly increased for both groups across arguments but there were differences in how students discussed reasoning in their arguments. In the methods section, I explained and identified four levels of evidence-model connections: Level 0: No connection between evidence and model, Level 1: Did not provide a reason for why the evidence and model were linked, Level 2: Provided a reason but the link between the evidence and model were superficial (did not justify how the evidence supports the model), and Level 3: Provided a reason and justification for why the evidence and model were linked. Figure 4.14 illustrates the evidence-model connections for all 120 arguments students wrote. Students in Group B had more arguments that included a level 0 evidence-model connection. They were less likely to provide a reason in their argument. Only three arguments for students in Group A included a level 1 evidence-model connection, and this was found at either the pretest or the beginning of the curriculum. More students in Group A had arguments that included evidence-model connections at levels 2 or 3. Level 2 evidence-model connections were found mostly in the middle of the curriculum. Level 3 evidence-model connections were often found for arguments towards the end of the curriculum or posttest.

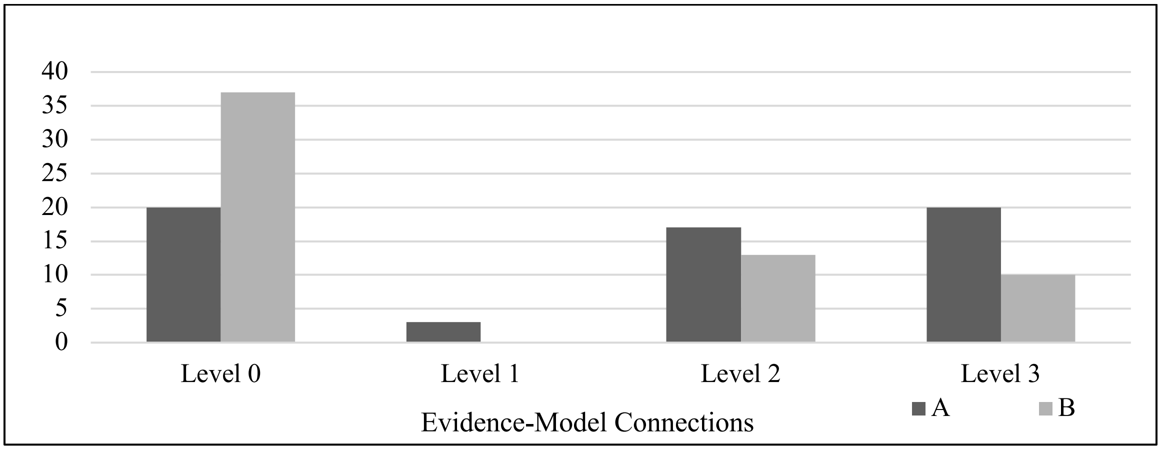


Figure 4.14. Evidence-model connections.

Figure 4.15 illustrates evidence-model connections across arguments. I found that students who had more than one argument with an evidence-model connection moved from level 0 to level 3 across arguments. There was a progression from not providing any reasoning, to providing a reason that did not justify how the evidence supported the model, to providing a reason with an explicit justification for why the evidence supported the model. For both groups, but especially Group A, evidence-model connections mostly increased from pretest to posttest with a decrease at the genetics assessment. In the next section, the analysis and findings for rebuttals are presented.

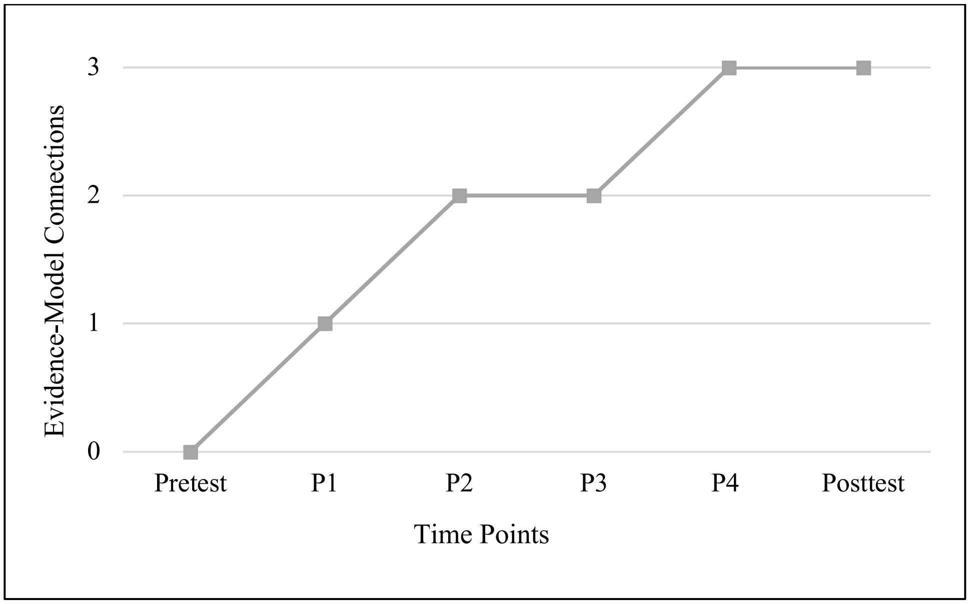


Figure 4.15. Progression of evidence-model connections.

Rebuttals

When students critique an argument, they compare and contrast models and evidence which leads to students discussing how evidence can either support or contradict the model. Arguments that include evidence that contradicts a model are important for science because such evidence is used to help rule out a model based on contradictory evidence. Rebuttals occur when a student says or implies that the evidence contradicts a model because the evidence supports one model and contradicts the other model or the evidence contradicts the model due to the quality of the evidence. Figure 4.16 illustrates rebuttals for the two groups. Out of 120 arguments, I identified 28 arguments with a rebuttal and most rebuttals were found in either the fifth or final argument with none in the first argument. Eighty-six percent of arguments with a rebuttal included a discussion of how evidence supported one model and used the higher quality evidence (evidence that supported the better supported model) to contradict the other model. For example, one student's argument within the evolution lesson ruled out the fighting model by using higher quality evidence that was an experiment and not anecdotal:

> Evidence that contradicts the fighting model is evidence 9. This evidence shows that body size didn't change much throughout the drought. The birds that survived had about the same body size as ones that died. The fighting model explains how birds with bigger bodies fight off the small birds which can't get to the food.

The other type of rebuttals were not found as often. Fourteen percent of the arguments stated that the evidence contradicted the model because the evidence was anecdotal (lower quality).

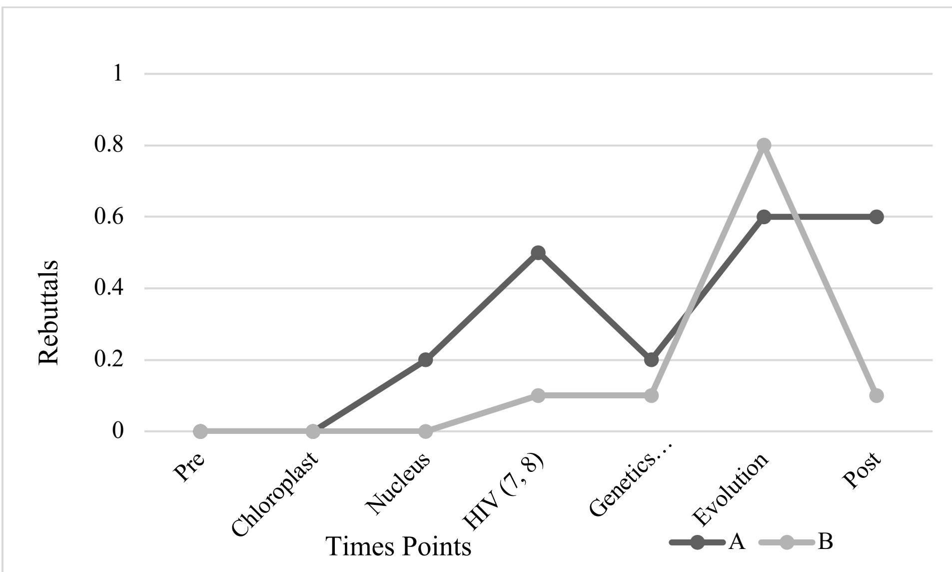
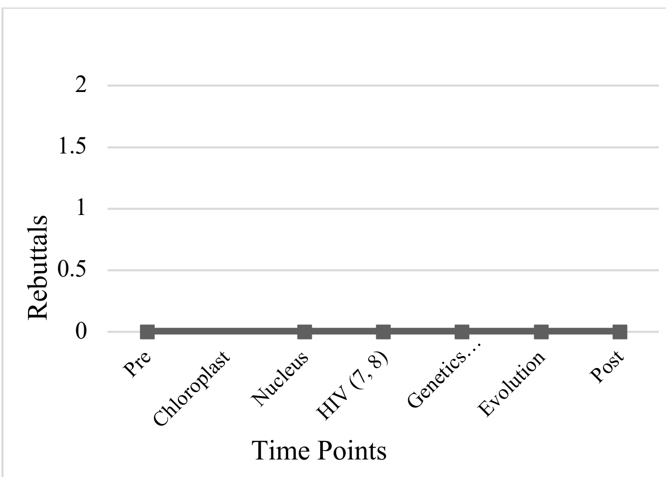


Figure 4.16. Rebuttals across arguments.

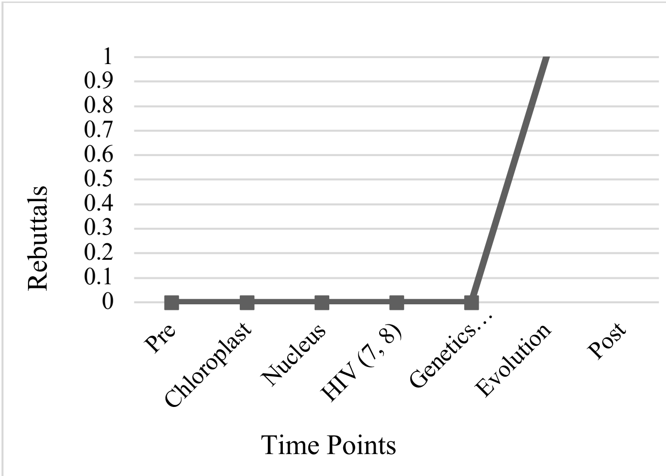
Students in Group A began to provide rebuttals earlier in the curriculum and provided more rebuttals as a group than Group B (31% of Group A, 16% of Group B). Six students in Group A included rebuttals in more than one argument. If a student in Group B included a rebuttal in his/her argument it was often for the evolution lesson (63% of Group A, 75% of Group B).

Figure 4.17a - d illustrates the different types of rebuttals. When looking at individual students' arguments across time, both groups exhibited similar patterns: (a) students did not provide a rebuttal (four students or 10% of students in Group A and 30% of students in Group B), and (b) students' rebuttals showed a small increase in number (six students or 40% of students in Group A and 20% of students in Group B). One additional pattern emerged for Group A: students' rebuttals increased, and then dropped at the genetics assessments or posttest (four students or 40% of Group A), and Group B: students only provided rebuttals for the evolution lesson (four students or 40% of Group B).

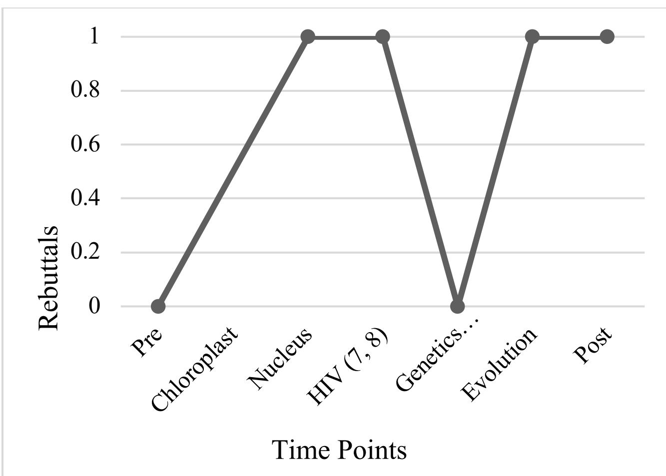
(a) Group A and B: No Rebuttals



(b) Group A and B: Small Increase



(c) Group A: Small Increase, Dropped at Assessment



(d) Group B: Evolution Lesson

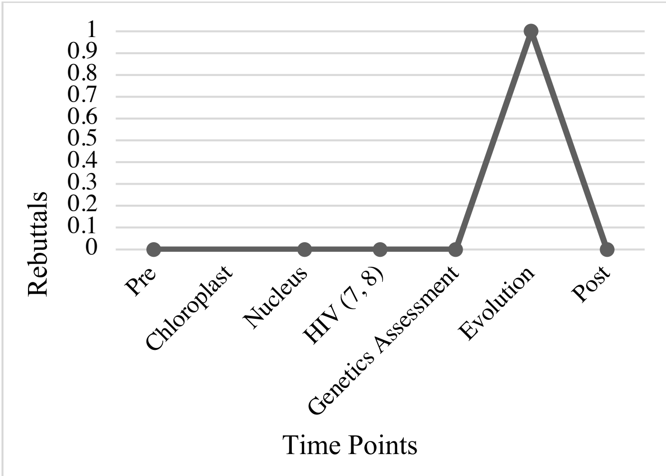


Figure 4.17a – d. Examples of the different types of rebuttals.

Do Certain Competencies Need to Be in Place Before Others?

In this section I discuss the associations between the different components of the argument (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals) to see if students' understanding and usage of one component was associated or dependent on another component, two components developed in parallel or there was no relationship between the two components. There are seven associations that I examined: (a) model choice and reasoning, (b) evidence usage and reasoning, (c) evidence conclusions and reasoning, (d) model choice and rebuttals, (e) evidence usage and rebuttals, (f) evidence conclusions and rebuttals, and (g) reasoning and rebuttals. Across the six arguments, I plotted the average score for each component in the matrix to identify the relationship between components.

Model Choice and Reasoning

Forty-four percent of students who choose the better supported model included reasoning in their written arguments. This suggested that students' inclusion of the better model was not linked to reasoning. However, I wanted to determine if there were group differences in the association between model choice and reasoning. Figure 4.18a and b plot the average score for model choice and reasoning for both Groups A and B across the six arguments. For both Groups A and B, model choice and reasoning developed slightly in parallel.

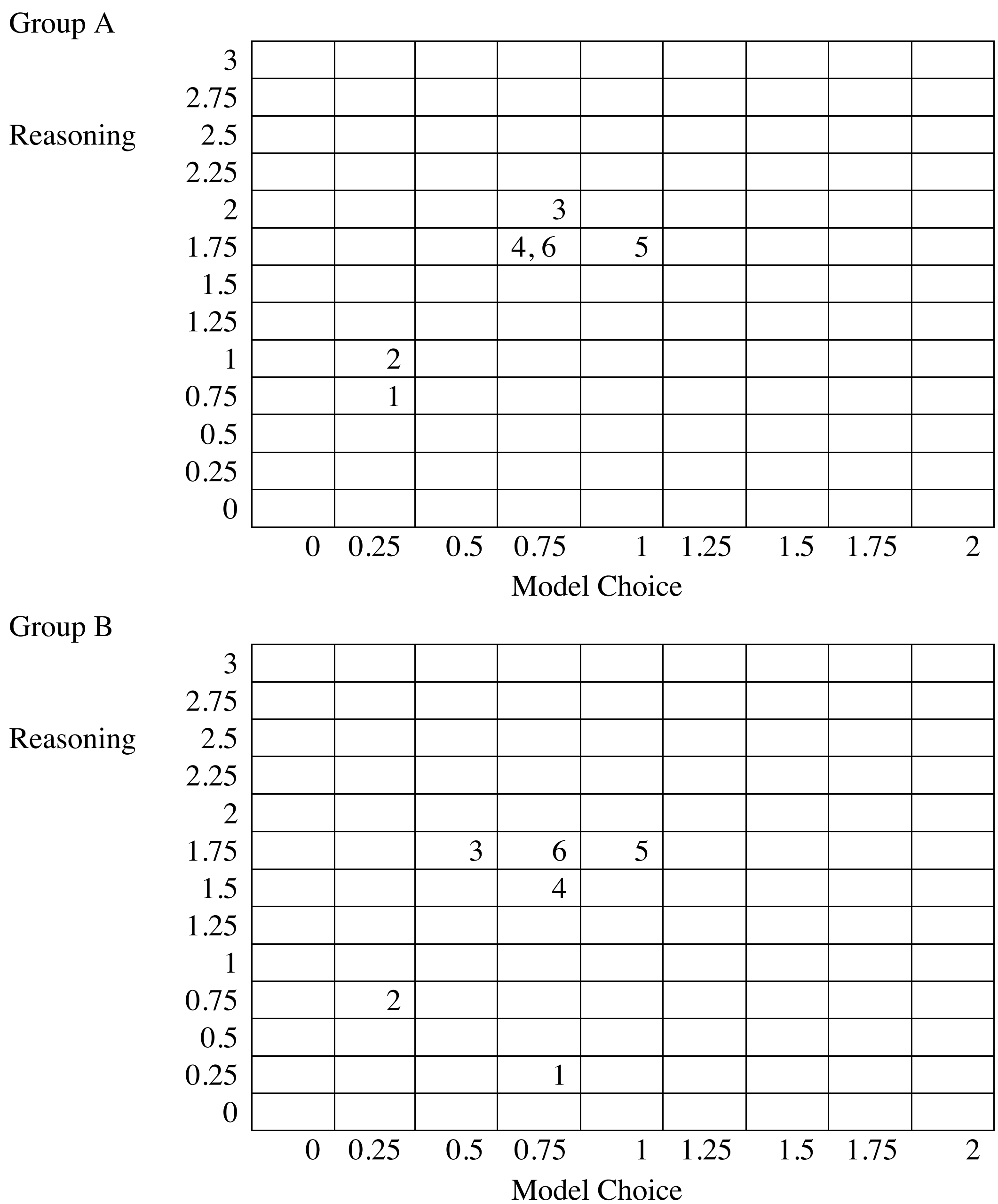


Figure 4.18 a and b. Model choice versus reasoning.

Evidence Usage and Reasoning

Figure 4.19a and b plots the average score for evidence usage and reasoning for both Groups A and B across the six arguments. For both Groups A and B, students mostly followed a parallel development between evidence usage and reasoning. For Group A, towards the end of the curriculum (time points 4 and 5), the relationship between evidence usage and reasoning changed with reasoning being dependent on evidence usage.

Next, I analyzed individual arguments' association between evidence usage and reasoning. I found that for both groups, reasoning was most often found in written arguments that included a higher level of evidence usage. Arguments needed to be at least a level 3 or 4 for evidence usage to be at a level 2 or 3 for reasoning. This was the case for arguments in Group A (66% of arguments that included evidence usage and reasoning) and Group B (52% of arguments that included evidence usage and reasoning).

Group A

Reasoning																	
3																5	
2.8																	
2.5																	
2.3																	
2													3				
1.8												6				4	
1.5																	
1.3																	
1										2							
0.8					1												
0.5																	
0.3																	
0																	
	0	0.3	0.5	0.8	1	1.3	1.5	1.8	2	2.3	2.5	2.8	3	3.3	3.5	3.8	4

Evidence Usage

Group B

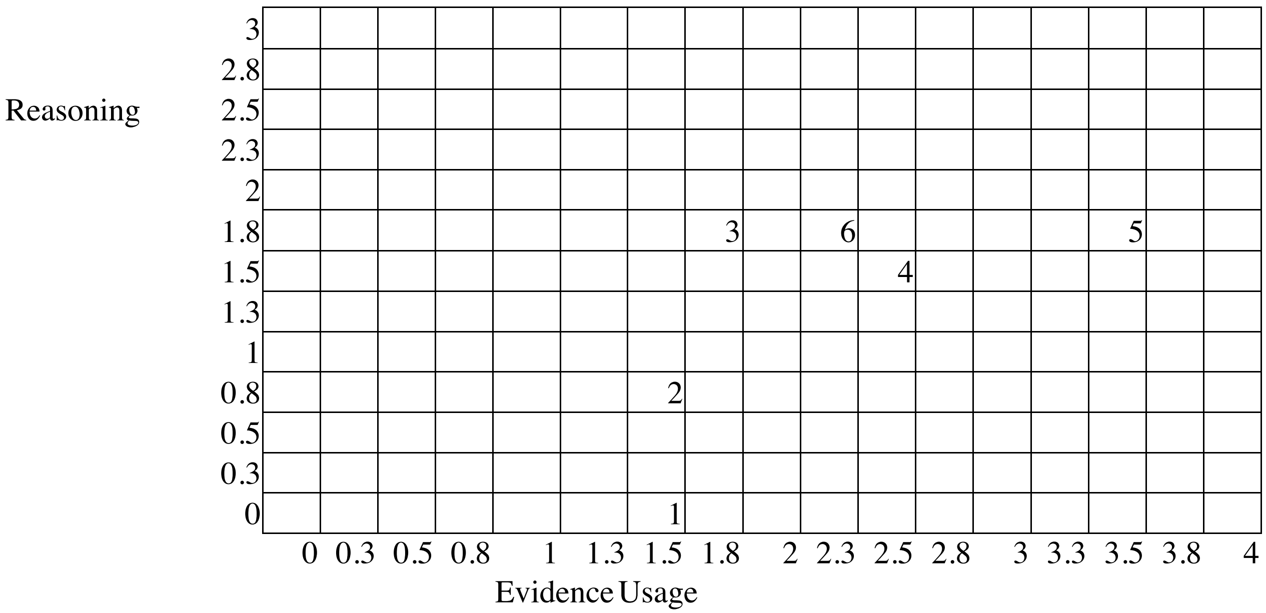


Figure 4.19a and b. Evidence usage versus reasoning.

Evidence Conclusions and Reasoning

For the association between evidence conclusions and reasoning, students who provided evidence conclusions provided a reason 48% of the time. One student wrote:

> I believe that the beak size model is the better model…One piece of evidence that supports it is evidence 5. It supports it because it shows that the weight actually gets lighter as the drought goes on and the beak size gets bigger. This evidence both supports the beak size model and contradicts the fighting model.

In this example the student provided evidence conclusions, described evidence and linked the evidence to the model and explained why they are linked (a higher level of reasoning). For both groups, students inclusion of evidence conclusions was linked to students inclusion of evidence usage and reasoning. Figure 4.20a and b plot average score for evidence conclusions and reasoning for both Groups A and B across the six arguments. For Group A, most arguments from pretest until the end of the curriculum (time points 1-5) developed in parallel between evidence conclusions and reasoning. For Group B, there was slight parallel development between evidence conclusions and reasoning.

Group A

	3											5		
	2.75													
Reasoning	2.5													
	2.25													
	2							3						
	1.75									4	6			
	1.5													
	1.25													
	1				2									
	0.75			1										
	0.5													
	0.25													
	0													
		0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3

Evidence Conclusions

Group B

Reasoning	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
3													
2.75													
2.5													
2.25													
2													
1.75						3	6	5					
1.5							4						
1.25													
1													
0.75			2										
0.5													
0.25					1								
0													

Evidence Conclusions

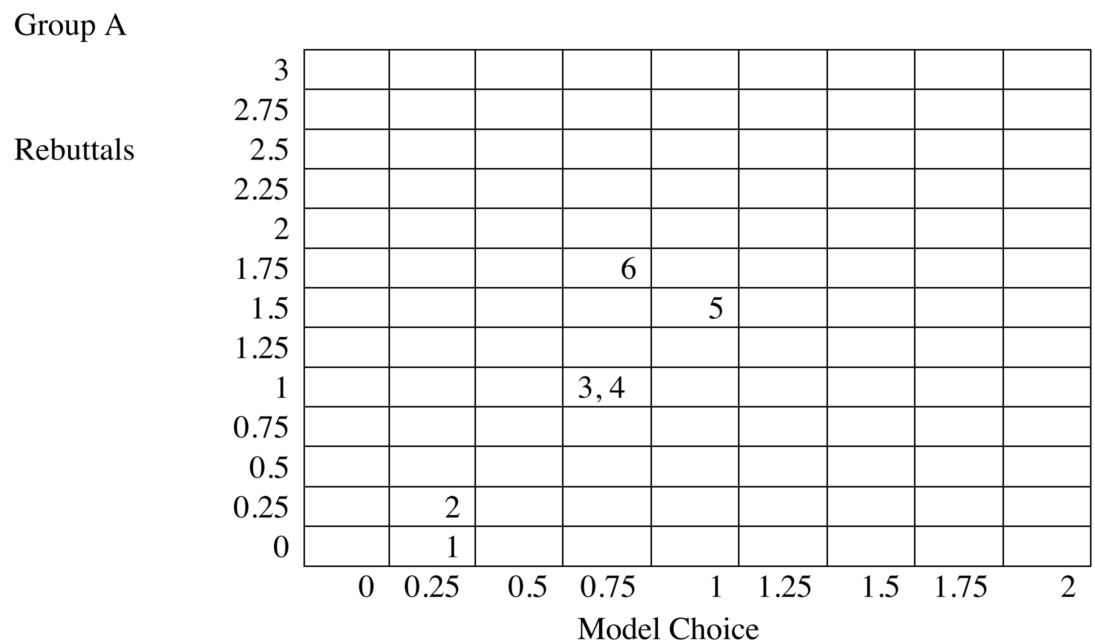
Figure 4.20a and b. Evidence conclusions and reasoning.

Next, I analyzed individual arguments' association between evidence conclusions and reasoning. I found that for both groups evidence conclusions were most often found in written arguments that included a higher level of reasoning. Arguments needed to be at least at level 2 for reasoning to be at level 1 for evidence conclusions. This was found for arguments in both Group A (62% of arguments that included evidence conclusions and reasoning) and Group B (61% of arguments that included evidence usage and reasoning).

Both Groups A and B's associations between model choice, evidence usage, and evidence conclusions versus reasoning developed in parallel. For individual arguments in both Groups A and B, there was a slight difference in associations between evidence usage and reasoning, and evidence conclusions and reasoning. Evidence usage needed to be at a higher level (level 3 or 4) for reasoning (level 2 or 3) to occur. However, the required levels differed for evidence conclusions. Reasoning needed to be at a higher level (level 2) than evidence conclusions (level 1).

Model Choice and Rebuttals

Figure 4.21a and b plot the average model choice and rebuttals for both groups across the six arguments. For both Groups A and B, model choice and rebuttals developed slightly in parallel and grew at the same time across most arguments. However, for individual arguments the groups differed. For Group A, half of the arguments where the student chose the better model, the argument also included rebuttals. Only 22% of written arguments in Group B where the student picked the better model the argument also included rebuttals.



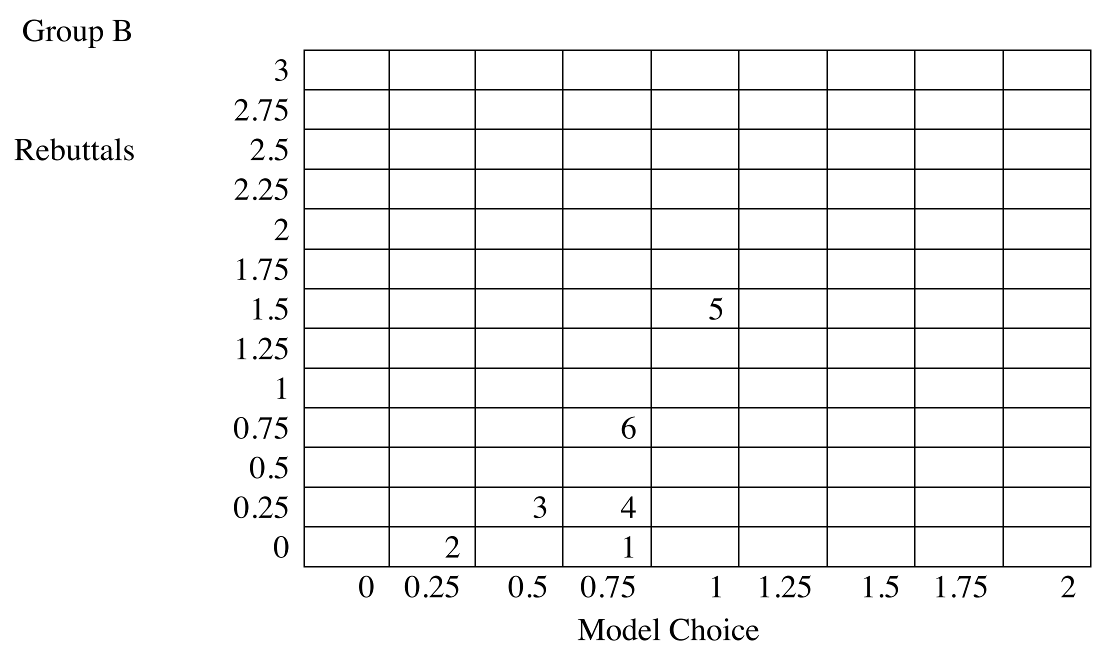


Figure 4.21a and b. Model choice versus rebuttals.

Evidence Usage and Rebuttals

Figure 4.22a and b plot the average evidence usage and rebuttals for both groups across the six arguments. The association between evidence usage and rebuttals depended on the point in the curriculum the argument occurred. For Group A, arguments written towards the end of the curriculum (time points 4 and 5), rebuttals were dependent on evidence usage. However, for arguments written at the beginning and middle of the curriculum and posttest (time points 1,2, 3, and 6) the associations between evidence usage and rebuttals developed slightly in parallel. For Group B, the association between evidence usage and rebuttals developed in parallel.

For individual arguments' association between evidence usage and rebuttals, I found that for both groups rebuttals were most often found in written arguments that included a higher level of evidence usage. Arguments needed to be at least at a level 4 or 5 for evidence usage to be at a level 2 or 3 for rebuttals. This was the case for arguments generated by Group A (65% of arguments that included evidence usage and rebuttals) and Group B (64% of arguments that included evidence usage and rebuttals).

Group A

Rebuttals \ Evidence Usage	0	0.3	0.5	0.8	1	1.3	1.5	1.8	2	2.3	2.5	2.8	3	3.3	3.5	3.8	4
3																	
2.75																	
2.5																	
2.25																	
2																	
1.75												6					
1.5																5	
1.25																	
1													3			4	
0.75																	
0.5																	
0.25										2							
0					1												

Evidence Usage

Group B

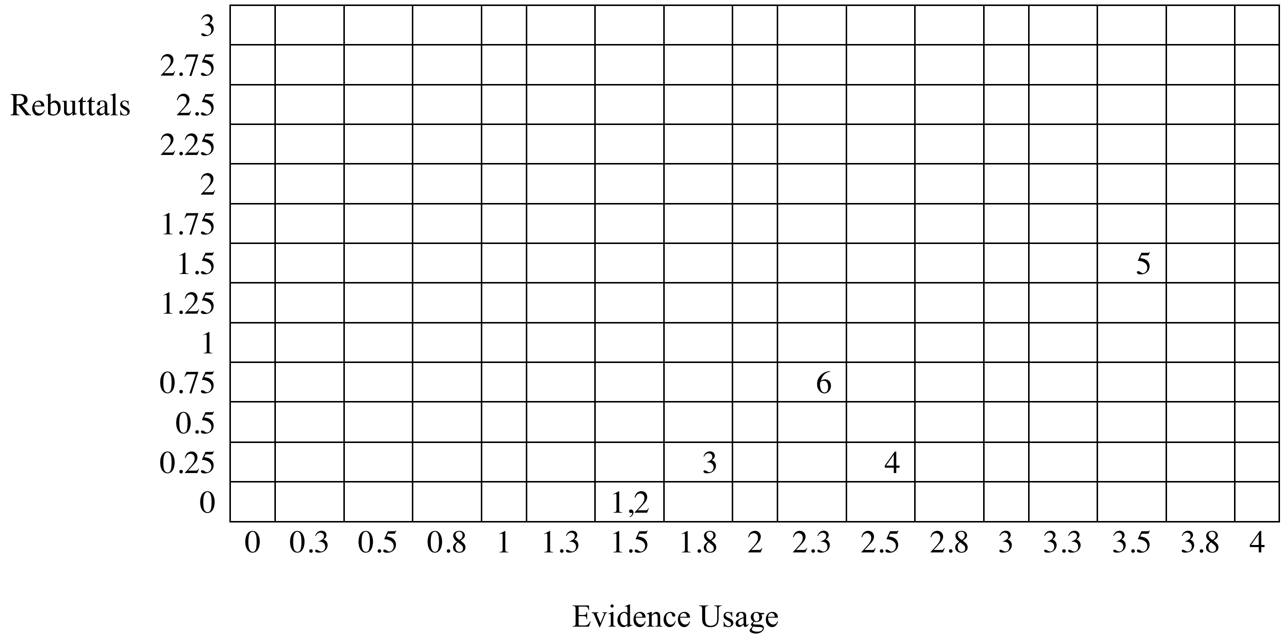


Figure 4.22a and b. Evidence usage versus rebuttals.

Evidence Conclusions and Rebuttals

Figure 4.23a and b plot the average evidence conclusions and rebuttals for both groups across the six arguments. Group A's inclusion of evidence conclusions was weakly linked to their inclusion of rebuttals with 30% of students providing evidence conclusions also providing rebuttals. For Group A evidence conclusions and rebuttals developed in parallel. For Group B, the association between evidence conclusions and rebuttals developed slightly in parallel.

For students' individual arguments, I found that both groups' rebuttals were most often found in written arguments that included a lower level of evidence conclusions. For arguments to be at a level 3 for rebuttals they needed to score at least at level 1 or 2 for evidence conclusions. This was the case for arguments in both Group A (56% of arguments that included evidence usage and rebuttals) and Group B (38% of arguments that included evidence usage and rebuttals).

Group A

Rebuttals	0	0.3	0.5	0.8	1	1.3	1.5	1.8	2	2.3	2.5	2.8	3	3.3	3.5	3.8	4
3																	
2.8																	
2.5																	
2.3																	
2																	
1.8										6							
1.5											5						
1.3																	
1							3		4								
0.8																	
0.5																	
0.3				2													
0			1														

Evidence Conclusions

Group B

Rebuttals	0	0.3	0.5	0.8	1	1.3	1.5	1.8	2	2.3	2.5	2.8	3	3.3	3.5	3.8	4
3																	
2.8																	
2.5																	
2.3																	
2																	
1.8																	
1.5								5									
1.3																	
1																	
0.8							6										
0.5																	
0.3						3	4										
0			2		1												

Evidence Conclusions

Figure 4.23a and b. Evidence conclusions versus rebuttals.

Reasoning and Rebuttals

Figure 4.24a and b plot the average reasoning and rebuttals for both groups across all six arguments. For both Groups A and B, reasoning and rebuttals developed slightly in parallel. For individual arguments there was no association between reasoning and rebuttals.

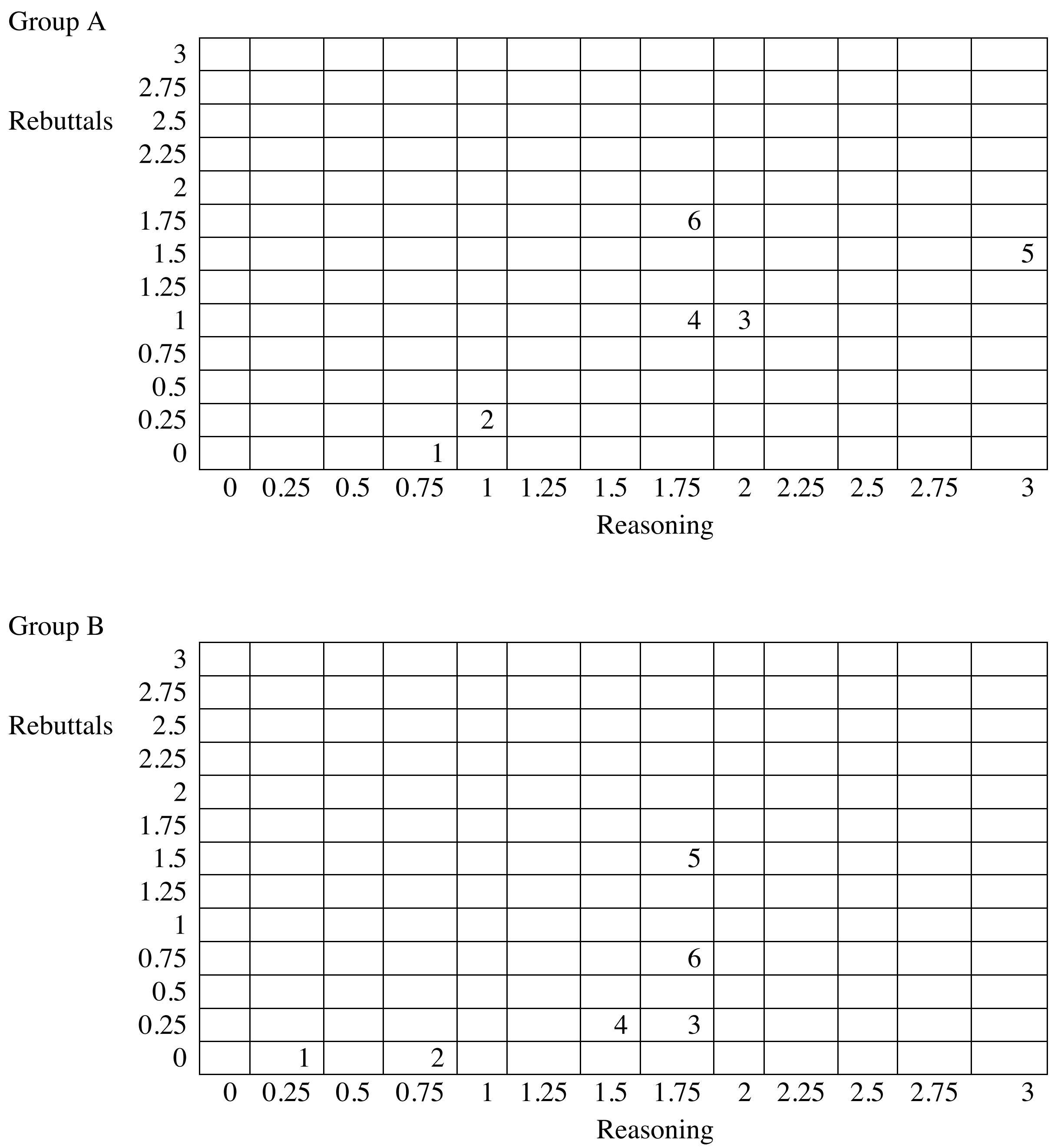


Figure 4.24a and b. Reasoning versus rebuttals.

For both Groups A and B, most associations developed in parallel. For individual arguments in both groups there was a slight difference in associations between evidence usage and rebuttals, evidence conclusions and rebuttals and reasoning and rebuttals. Evidence usage needed to be at a higher level (level 4 or 5) for rebuttals (level 2 or 3) to occur, but it differed for evidence conclusions. Rebuttals needed to be at a higher level (level 3) than evidence conclusions (level 1 or 2).

The Value of Supports Embedded in the Curriculum

Although Groups A and B experienced the same lessons and assessments, there were differences between groups in their written arguments. Particularly, in the components of the arguments, especially evidence usage, evidence conclusions, reasoning and rebuttals and associations between components. Osborne and colleagues (2016) and McNeill (2011) suggest that the quality of students' written argument is determined by students' understanding of content, but I did not find this. I found that both groups had a similar understanding of content, models and evidence. Any difference in understanding was mostly ascribed to students in Group B who demonstrated a better understanding of content and certain pieces of evidence. Although both groups understood the content, models and evidence, there were differences between groups in their evidence usage. Particularly including higher quality evidence in their written arguments, reasoning and rebuttals.

It was important for students' arguments to go beyond following a structure and move into using elements to construct higher quality arguments. Use of elements was found in some students' arguments as they continued across the curriculum. By ruling out that there was no difference in both groups' understanding of content, models and evidence, I next evaluated other possibilities for differences between groups for the quality of written arguments. Across the curriculum, students had multiple opportunities to evaluate the quality of the evidence and discuss model goodness. Evaluation of evidence occurred through the use of supports that were introduced at different points of the curriculum and were useful for students as they wrote arguments at the end of the lessons and assessments.

Supports embedded in the curriculum were valuable for students, and McNeill (2011) found that students struggled with identifying the best or most appropriate evidence to support the claim if they were responding to an open-ended question with less scaffolding. In the methods section, I explained the scaffolds that were included in the PRACCIS curriculum. The supports included the model-evidence link (MEL) matrix, evidence ratings, class criteria for model goodness and argumentation rubric.

The MEL matrix, evidence ratings and the class criteria list for model goodness were introduced earlier in the curriculum than the argumentation rubric. For the argumentation rubric to be useful for students, students needed to be able to evaluate the models and evidence and examine the relationship between models and evidence, so the MEL matrix was introduced first. The development of the class criteria list was the first lesson students completed and as students were introduced to more content throughout the curriculum they were able to go back to revisit and revise this list (Pluta et al., 2011; Rinehart et al., 2014). The purpose of the criteria last was for students' to use the list evaluate the goodness of both models.

The MEL matrix supported students as they evaluated competing models and multiple pieces of evidence by asking students to decide how each piece of evidence supported, contradicted, or was irrelevant to each model. The main purpose of this scaffold was to support students' thinking about the relationship between the model and pieces of evidence (Rinehart et al., 2014). The MEL matrix also included an evidence quality-rating component. Figure 4.25 illustrates an example of the MEL matrix. By using the MEL to rate the quality of the pieces of evidence students were more likely to produce higher quality arguments – especially in the case of Group A. Arguments written by students in Group A were more likely to provide a correct conclusion for a piece of evidence, describe evidence in their arguments and use higher quality evidence to support the model chosen.

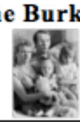
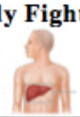
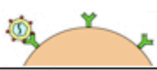
Evidence Quality Rating	**Model 1** **ATTACK-AND-DESTROY**	**Model 2** **HIV-CAN'T-GET-IN**
#1. The Burke Family.  ☐		
#2. How the Body Fights the Hepatitis Virus.  ☐		
#3. Jen's Favorite Genes.  ☐		
#4. CCR5 Research.  ☐		
#5. DNA Analysis  ☐		

Figure 4.25. MEL matrix.

For lessons that included the MEL matrix (all lessons except the chloroplast lesson), students in both groups included model-evidence relationships in their written arguments (79% of arguments for Group A and 64% of arguments for Group B). For both groups, the MEL matrix was least used for the nucleus lesson but after this lesson both groups increased in using the MEL matrix (100% for Group A and 72% for Group B). Both groups included evidence that supported the model chosen. On occasion, use of the MEL matrix led to the inclusion of reasons and rebuttals in students' written arguments.

For reasoning, students in Group A were more likely to include a reason in their argument than students in Group B (67% for Group A and 38% for Group B). For lessons that included the MEL matrix, both Groups A and B included the model-evidence relationship identified in the matrix in almost all of the arguments that included a reason (91% of arguments for Group A and 88% of arguments for Group B). One difference between Groups A and B was that Group A provided more reasons in both the lessons that included the MEL matrix (Group A: 78%, Group B: 57%) and assessments that did not include the MEL matrix (Group A: 56%, Group B: 22%).

For rebuttals, only 28 arguments (16 rebuttals in lessons and 12 in assessments) included a rebuttal, but the MEL matrix was helpful for students. For instance, a student that included a rebuttal at the nucleus lesson used the MEL matrix to support his/her written argument and continued to use the MEL matrix for rest of the arguments written across the curriculum. For the different lessons that used the MEL matrix, out of the 16 arguments that included a rebuttal, 14 arguments used the MEL matrix to support their discussion of how the evidence contradicted one of the models. Students in Group A were slightly more likely to include a rebuttal and often used the MEL matrix to support their rebuttals (58% for Group A and 43% of Group B). The rest of the arguments that included a rebuttal were mostly found at the posttest and written by students in Group A (six students for Group A and three students for Group B). Although more students in Group A included reasoning and rebuttals at the posttest, most students in Group B and some in Group A may have benefitted from completing the MEL matrix, using the other supports and discussing the models and evidence before writing their arguments.

How Did the Introduction of the Argumentation Rubric Change Students' Arguments?

Table 4.4 illustrates the argumentation rubric. It was the final support introduced to students. The argumentation rubric was introduced in the middle of the curriculum (HIV lessons 7 and 8). Students used the rubric for the evolution lesson but not for the assessments (genetics assessment and posttest). The rubric reminded students of the components of the argument that needed to be addressed to generate high quality arguments.

Table 4.4

Evidence-based Argumentation Writing Rubric

	0	**1**	**2**	**3**
Number of evidence discussed	No evidence mentioned	Only one evidence discussed	More than one evidence discussed	ALL relevant evidence is discussed
Writing about evidence	No evidence mentioned	Mentions evidence	Mentions and describes evidence	Mentions, describes AND explains how evidence supports the model
Counter evidence	No mention of evidence that contradicts other model	Mentions evidence that contradicts another model	Mentions and describes evidence that contradicts another model	Mentions and describes evidence AND explains that contradicts another model

For Groups A and B, I examined components of argument addressed in the argumentation rubric to determine if differences existed between arguments written before the rubric was introduced (the nucleus lesson) and arguments written after the rubric was introduced (HIV lessons 7 and 8). Both Groups A and B exhibited differences between the written arguments before and after the rubric was introduced with Group A's arguments showing greatest improvement. The differences between written arguments for Group A were found for each component: (a) model choice (the nucleus lesson: 30%, the HIV lessons: 90%), (b) evidence usage (the nucleus lesson: 50%, the HIV lessons: 100%), (c) evidence conclusions (the nucleus lesson: 30%, the HIV lessons: 100%), (d) reasoning (the nucleus lesson: 50%, HIV lessons: 100%), and (e) rebuttals (the nucleus lesson: 10%, the HIV lessons: 50%). Group B also demonstrated differences between written arguments for each component but had smaller gains before and after the argumentation rubric was introduced: (a) model choice (the nucleus lesson: 30%, the HIV lessons: 60%), (b) evidence usage (the nucleus lesson: 60%, the HIV lessons: 100%), (c) evidence conclusions (the nucleus lesson: 30%, the HIV lessons: 70%), (d) reasoning (the nucleus lesson: 30%, HIV lessons: 80%), and (e) rebuttals the nucleus lesson: 0%, the HIV lessons: 10%).

After the introduction of the argumentation rubric, students' written arguments from the HIV lesson differed from those arguments generated before the rubric was introduced (i.e., the nucleus lesson). Both groups improved in choosing the better model, evidence usage, evidence conclusions, reasoning, and rebuttals. It is possible that the introduction of the argumentation rubric helped students to write higher quality arguments, and it is a valuable support that worked well in conjunction with the other supports (e.g., evidence rating, criteria lists, MEL matrix). Also, I found that when the supports were removed for the genetics assessment and posttest there was a decline in students' written arguments, especially for Group B at the posttest. For Group B there was a decline in reasoning and the number of rebuttals even though both were found in other arguments written during the curriculum. This suggests that supports were valuable for students and were still needed to produce high quality arguments, especially for students in Group B until the posttest.

Why Was There a Drop at the Genetics Assessment?

The final section of this chapter focuses on the genetics assessment. The analysis of the composite argument score and the various components of the argument (model choice, evidence usage, evidence conclusions, reasoning, rebuttals) followed a pattern that was found for both groups — a decline at the genetics assessment. Figure 4.26 illustrates the differences in students' evidence usage across the lessons that occurred before and after the genetics assessment. For reasoning, there was a large decrease in both groups, with Group A doing slightly better on the genetics assessment. For Group A, students provided more reasons than students in Group B before and after the genetics assessment. Figure 4.27 illustrates the differences in reasoning across the lessons that occurred before and after the genetics assessment. For Group A, most arguments from pretest until the end of the curriculum (time points 1-5) demonstrated parallel development between evidence conclusions and reasoning. For Group B, there was slight parallel development between evidence conclusions and reasoning. Figure 4.28 illustrates the differences in rebuttals across lessons that occurred before and after the genetics assessments.

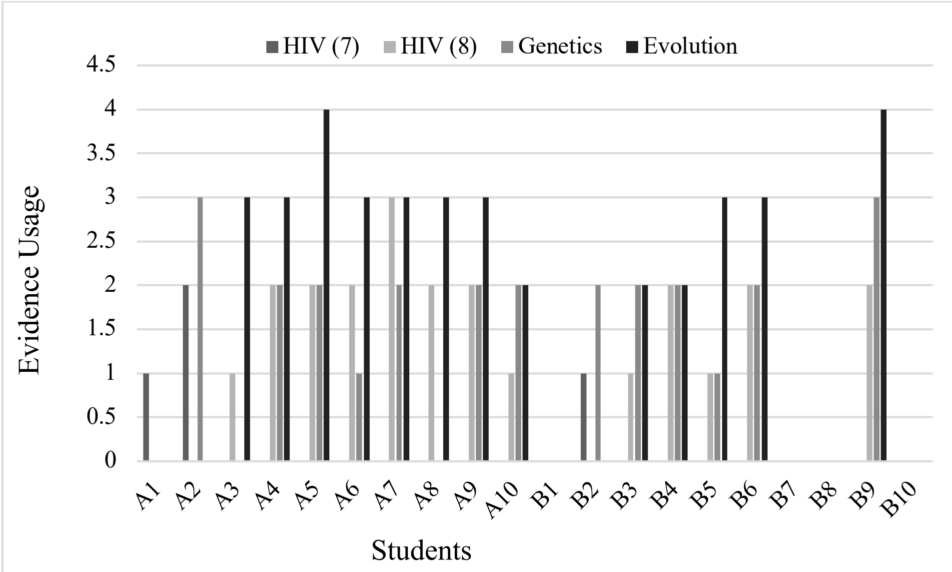


Figure 4.26. Evidence usage before and after the genetics assessment.

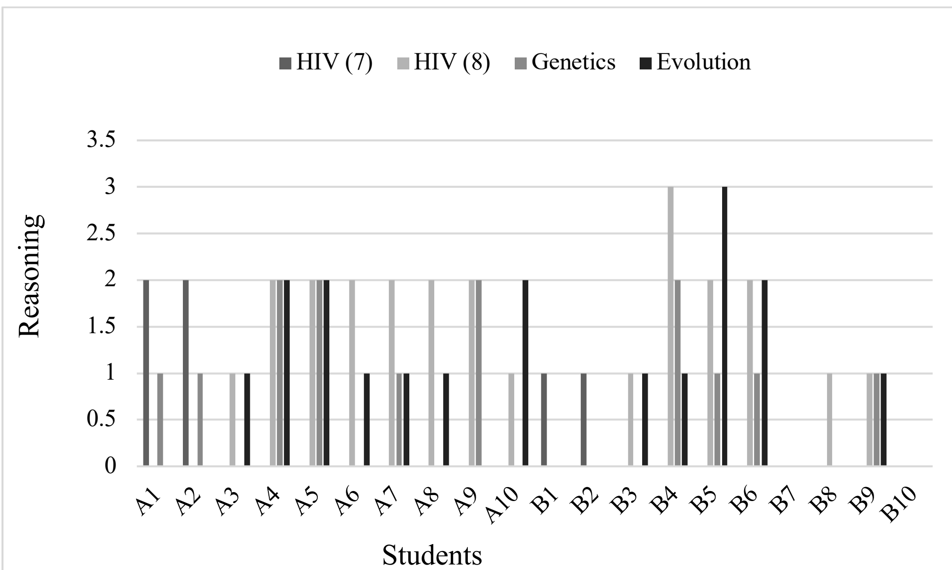


Figure 4.27. Students' reasoning for HIV lessons, genetics assessment, and evolution lesson.

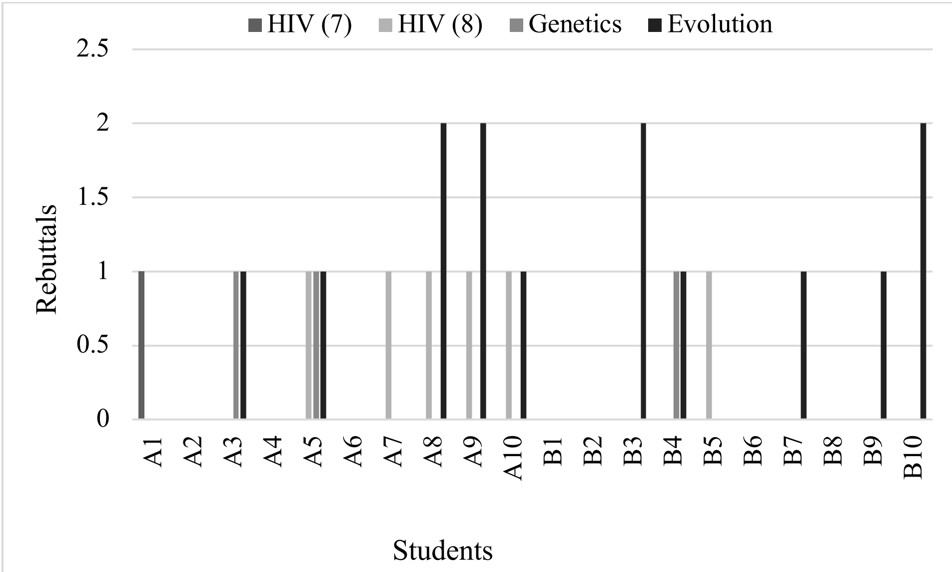


Figure 4.28. Students' rebuttals for HIV lessons, genetics assessment, and evolution lesson.

Although there were differences in students' arguments across the curriculum, the findings identified some patterns that help explain why and how change occurred. For instance, the decline in reasoning, evidence usage, and rebuttals for the genetics assessment can be attributed to classroom instruction and the removal of scaffolds. For all writing prompts that occurred during the curriculum, students participated in whole group discussions where the teacher reviewed the models and pieces of evidence. This provided students, who may have had difficulty understanding the models and evidence initially, an opportunity to review the models and evidence as a class. Resulting arguments were higher in quality. For the pretest, posttest, and genetics assessment, students were not provided with the opportunity to review the models and evidence as a class or have access to most of the scaffolds (e.g., argumentation rubric, MEL matrix, evidence ratings). This suggests why there was a decrease at the genetics assessments and sometimes at posttest, and that the abrupt removal of scaffolds was detrimental. Students in both groups still needed the scaffolds, especially Group B, at the posttest.

Chapter 5

Discussion

In this study I examined changes in students' arguments across multiple components over six written arguments (pretest, four different time points in the curriculum and posttest). Students' capacity to generate high quality arguments improved as a result of having multiple opportunities to generate arguments throughout the curriculum. I wanted to determine how students' written arguments changed across a curriculum and if there were differences in students' capacity to argue. I analyzed students' written arguments at different time points and their understanding of models, content, and evidence. By examining two groups that differ in gains made from pretest to posttest (Group A: greater gains from pretest to posttest, Group B: lesser gains from pretest to posttest), I expected to find differences such that Group B's arguments did not change as much from pretest to posttest (even though they participated in the same instruction). I expected that differences in arguments between Groups A and B occurred because of differences in their understanding of content, models, and evidence (e.g., McNeill, 2011; Osborne et al., 2016). However, I found that both groups had a similar understanding of content, models, and evidence even though there were differences in their written arguments. Another reason for differences between groups in their written arguments included which evidence students used in their arguments and their use of supports embedded in the curriculum. I did find that students' usage of evidence differed between groups with students in Group A including more high quality evidence in their arguments – often accompanied by reasoning.

For both groups, I also expected that certain competencies needed to be in place before others developed, and that students needed to be supported throughout the curriculum in order for their written arguments to demonstrate continued improvement. For instance, before a student included a reason in his/her argument the student needed to choose a model and use evidence to explain why he/she picked the model. As part of the curriculum, students were introduced to a suite of scaffolds that were somewhat useful in supporting students as they evaluated models and evidence to help improve students' written arguments. As students' written arguments improved, I wanted to determine if they showed gradual improvement in argument writing or their arguments change suddenly. Results from this study could be used to improve previously designed argumentation learning progressions (Berland & McNeill, 2010; Osborne et al., 2016) by suggesting modifications that better support students' capacity to generate high quality arguments.

The intention of the argumentation learning progressions was to establish a sequence for how arguments change over time to become more sophisticated. However, both learning progressions evaluate arguments over years and do not focus on change in a more nuanced way (Berland & McNeill, 2010; Osborne et al., 2016). For this study, I examined arguments across the same school year to get a better understanding of what progression students follow towards competency in scientific argumentation and mastery of specific scientific content. This approach determined if there were differences between the two groups. The learning progressions hypothesized how argumentation developed, the value of science content knowledge and ways to support student learning (Berland & McNeill, 2010; Osborne et al., 2016). The research questions for this study are: (1) What are the trajectories of change over time in students' written arguments? (a) What are the differences in the quality and structure of students' arguments? (2) In written argumentation, do certain competencies need to appear before others develop? If yes, what are the competencies? The research questions help address modifications for the learning progressions. In particular what components of the argument develop first and the association between components. In the upcoming sections, each research question is addressed and follows with a discussion for implications based on the findings. The final section addresses limitations of this study.

What are the Trajectories of Change Over Time in Students' Written Arguments?

To understand the differences between Groups A and B in terms of how students' arguments changed over time, I began by analyzing Group A and B's whole arguments across the six essays per student. Students were selected for Group A because they had greater gains from pretest to posttest in total score, and students in Group B were selected because they made lesser gains from pretest to posttest. Due to the small sample size, I did not run any tests for statistical significance. To examine differences between groups I am referring to literal differences found through plotting the composite argument score and components of the arguments. A composite argument score for each argument was calculated using the components of the whole argument that included: model choice, evidence usage, evidence conclusions, reasoning, and rebuttals. Group A's composite argument score mostly increased across lessons and assessments, with a decrease at the genetics assessment. Group B's had a slower increase, with a drop off at the posttest.

For content and evidence understanding, Group B numerically outperformed Group A (especially earlier in the curriculum). Both groups had similar understanding of the models. This finding was surprising because Group A showed greater improvement across arguments for the whole argument versus Group B. Previous studies in argumentation suggest that the quality of students' written argument is determined by students' understanding of content (e.g., McNeill, 2011; Osborne et al., 2016). To unpack the differences between groups in their written arguments, I analyzed each component (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals) of the argument by plotting for both the groups' and individual students' arguments from pretest to posttest. Groups differed when the pattern the groups followed differed or when one group showed a greater increase (for instance, an increase in reasoning across arguments) than the other group. I also looked for patterns within each group to determine if patterns existed between groups. Groups A and B had similar scores for model choice across arguments, but for the other components of the argument the scores differed between groups. The next two components, evidence usage and evidence conclusions, focused on students' understanding of the main points of the evidence. Although both groups understood the evidence, they differed in how they used the evidence and if they provided evidence conclusions. By the end of the curriculum (posttest), most students in Group A provided evidence conclusions and described at least one piece of evidence. Whereas by the posttest, only half of the students in Group B described at least one piece of evidence. This suggested that students needed not only to understand the evidence, which both groups did, but also to engage actively with each piece of evidence, which Group A did more extensively.

Reasoning was the next component analyzed. Group A made greater gains for reasoning across arguments than Group B. Students in Group A were more likely to include reasoning in their arguments for the HIV lessons, the evolution lesson and posttest. Group B had a smaller increase in reasoning, with a drop off at the posttest. Between Groups A and B, for those students who engaged in reasoning, their reasoning became more nuanced by the end of the curriculum. Their reasoning moved from not explicitly justifying how the evidence supported a model to providing explicit justification for why the evidence supported the model.

Formulating rebuttals was the most complex component of the argument because rebuttals asked students to evaluate models and evidence, and identify evidence that contradicted one of the models. Rebuttals provide evidence and reasoning to rule out a possible claim, or why alternative claims are unsuitable (Krajcik & McNeill, 2015). Most of the arguments written by students in Groups A and B did not include rebuttals (only 28 arguments out of 120 arguments included rebuttals). However, I found that students in Group A were more likely to provide rebuttals and they started to provide rebuttals earlier in the curriculum than Group B. Most students in Group B started to include rebuttals in their arguments for the evolution lesson. Although Group A's arguments increased in the use of most components (evidence usage, evidence conclusions, reasoning, rebuttals) of more so than Group B, one similarity between groups was that they demonstrated a decrease at the genetics assessment. I discuss the genetics assessment in greater detail later in this chapter. In the next section, I discuss the differences in the quality of students' arguments, using components of the argument and additional analyses to explain why the quality of arguments differed even though students' understanding of was similar.

What Are the Differences in the Quality and Structure of Students' Arguments?

In this study, I define structure of argument as claim, evidence and reasoning. Argument quality is defined as following structural accuracy and the inclusion of evidence that supported the model chosen (especially the higher quality evidence). From pretest to posttest, I found differences in the quality and structure of students' arguments. The initial arguments written after the pretest were mostly low quality and structure. Students did not use evidence in support of the model they were discussing and rarely made a connection between the model and evidence. By the middle of the curriculum (the HIV lessons), the quality of students' arguments started to improve, especially for Group A. There were two suggestions for the change in students' written arguments in the middle of the curriculum: (a) students were introduced to the argumentation rubric and (b) students used higher quality evidence to support the model chosen.

As part of the curriculum students were provided with two competing models and evidence that varied in quality. Higher quality evidence supported the better model or supported the better model and contradicted the other model. The shift in argument quality included students choosing the better-supported model, using criteria such as a "good model explains the evidence" and using higher quality evidence to support the model chosen. I found that students in both groups understood the models, used similar amounts of epistemic criteria to evaluate model goodness and were more likely to pick the better-supported model in their written argument. However, students differed in how they discussed the model chosen. Group A was more likely to use higher quality evidence in their arguments and, across arguments, the usage of higher quality evidence increased for Group A and mostly stayed the same for Group B. This suggested another difference between groups for argument quality. It is possible that evidence understanding preceded evidence evaluation and evidence usage. Group B had a slightly better understanding of evidence, but Group A included higher quality evidence in their arguments. By the end of the curriculum, students in Group A had a good understanding of the evidence and were more likely to include multiple pieces of evidence in their arguments that they evaluated as good. When students rated evidence as good, they often included a detailed description of the evidence and often provided reasoning for why the evidence supported the model chosen.

The inclusion of reasoning and/or rebuttals in arguments suggested that the argument was higher in quality. For reasoning, students needed to be able to understand and evaluate both models and each piece of evidence, and required the ability to coordinate a claim and evidence (Sandoval, Sodian, Koerber, & Wong, 2014). Over time, students in Group A were more likely to provide reasons in their arguments and their reasoning increased in complexity (including explicit justification for why the evidence supported the model). For some students in Group B they still struggled with providing a reason (two students in Group B did not provide any reason in any argument, and most students in Group B did not provide a reason for the posttest). Rebuttals were not often included in arguments, and were only found in arguments starting in the middle of the curriculum.

Supports embedded in the curriculum (e.g., model criteria lists, MEL matrix, evidence ratings and argumentation rubric) and class discussions about the models and evidence were valuable for students and helped students write higher quality arguments. The model criteria lists and evidences ratings helped students evaluate the models and evidence. The MEL matrix was somewhat useful for helping students include reasoning and rebuttals in their written arguments. The argumentation rubric reminded students of the various components of the argument that needed to be addressed in order to write higher quality arguments. Group A showed greater gains for all components of the argument after the rubric was introduced.

For both groups, there was a drop off in reasoning and rebuttals at the genetics assessments and/or posttest. This suggested that supports need to be embedded into the curriculum at points when students are asked to evaluate model goodness (criteria lists), evaluate the quality of evidence (evidence ratings), make links between the models and evidence (MEL matrix), and generate arguments (argumentation rubric). Students may benefit from having these scaffolds in the assessments as well. The abrupt fading of supports prior to assessments was detrimental. Students still needed the supports, and perhaps more so for students in Group B.

Do Certain Competencies Need to be in Place Before Others Develop?

To examine if certain competencies needed to be in place before others develop, I examined associations between different components of the argument (model choice, evidence usage, evidence conclusions, reasoning, and rebuttals). I wanted to determine if students' understanding and usage of one component was associated or dependent on another component, two components developed in parallel or there was no relationship between the two components. Seven associations were examined: (a) model choice and reasoning, (b) evidence usage and reasoning, (c) evidence conclusions and reasoning, (d) model choice and rebuttals, (e) evidence usage and rebuttals, (f) evidence conclusions and rebuttals, and (g) reasoning, and rebuttals. From the analysis of written arguments across the curriculum, most of the associations for both groups developed in parallel.

For the first three associations: (a) model choice and reasoning, (b) evidence usage and reasoning, and (c) evidence conclusions and reasoning, both groups' associations mostly developed in parallel. For instance, when students included evidence conclusions in their arguments, they also included reasoning. I did find for individual arguments, evidence usage had to be at a higher level (describing multiple pieces of evidence) than reasoning (include one reason in the argument). However, for evidence conclusions and reasoning, reasoning needed to be at a higher level (provide justification for reasoning) than evidence conclusions (included one evidence conclusion in the argument). For arguments that included a reason, students needed to included multiple pieces of evidence and describe the evidence (a detailed description of the evidence content).

For the final four associations: (d) model choice and rebuttals, (e) evidence usage and rebuttals, (f) evidence conclusions and rebuttals, and (g) reasoning and rebuttals, both groups' associations between components mostly developed in parallel. For individual arguments, evidence usage (describing multiple pieces of evidence) had to be at a higher level than rebuttals (stating that the evidence contradicted the model). However, for evidence conclusions and rebuttals, rebuttals had to be at a higher level (stating that the evidence supports one model and contradicts the other model) than evidence conclusions (providing a conclusion about a piece of evidence). For an argument that included a rebuttal where the student wrote that the evidence supported one model and contradicted the other model, the argument included one evidence conclusion. For individual arguments there was no relationship between reasoning and rebuttals.

For both groups, reasoning and rebuttals developed in parallel with model choice, evidence usage and evidence conclusions. As students picked a model and described evidence it was likely that they included a reason or rebuttal. There were individual argument differences found for the associations of reasoning and rebuttal, evidence usage and evidence conclusions. Arguments needed to be at a higher level of evidence usage to include a reason or rebuttal, but for evidence conclusions, reasoning and rebuttals had to be at a higher level. For evidence conclusions, students needed to understand the main points of the evidence and be able to explain the findings in the evidence (if it was a study that students were learning about). Explaining findings may have been harder for students to do than just describing the evidence. It may be that choosing a model and describing evidence develops first in arguments before students include reasoning or rebuttals in their arguments. Also students begin to explain how evidence supports the model chosen before including evidence conclusions.

Implications for Instruction

Both argumentation learning progressions (Berland & McNeill, 2010; Osborne et al., 2016) hypothesized how argumentation progresses. For instance an argument moves from identifying a claim to identifying evidence to support the claim. However, neither progression discusses the evaluation of competing models and evidence that vary in quality. Including students' ability to evaluate models and evidence, their usage of evidence, and evidence conclusion to support the model chosen can identify change in the quality of an argument. I did find that most components of the argument developed in parallel for both groups, but for individual arguments students needed to describe evidence to be able to include a reason. Both learning progressions tend to focus on structure and suggest differences in arguments occur because of differences in students' understanding of scientific content. From my analysis, I found that Group B had a slightly better understanding of content, but Group A's arguments improved more than Group B. Students in Group A were more likely to use higher quality evidence, and include reasoning and rebuttals in their written arguments.

Due to my findings and evaluation of the argumentation learning progressions, I make suggestions to revise the instruction of this curriculum. The suggestions are: (a) include long term, persistent supports in the curriculum (Berland & McNeill, 2010), (b) include supports that help students as evaluate evidence because the ability to coordinate a claim and evidence requires various reasoning practices (Sandoval et al., 2014), and (c) develop criteria to promote argument quality (Chinn et al., 2014) and include epistemic criteria for evidence quality.

The three suggestions focus on the inclusion of supports and students' evaluation of evidence. The first suggestion is to not fade supports because I found that when supports were removed there was a decline in the quality of students' arguments. Although supports should decrease over time, students still need some supports through high school (Berland & McNeill, 2010). When the supports were removed in this study at the genetics assessment and posttest, both Groups A and B decreased in including evidence, reasoning and frequency of rebuttals in their written arguments at either the genetics assessment and/or posttest. The arguments students wrote when they had access to the supports suggest that both the MEL matrix and argumentation rubric were helpful in supporting students to include reasoning and rebuttals in their written arguments. Middle school students need to be supported if we want them to critique two competing arguments (Osborne et al., 2016). The MEL matrix was helpful because it allowed students to identify the relationship between the models and evidence.

The second suggestion focuses on providing a greater emphasis on evidence evaluation and the discussion of evidence quality. I found that when students were able to describe evidence they were more likely to include a reason (justify how the evidence supports the model chosen) in their written argument. Students need to understand that not all evidence is equal (Chinn et al., 2014). I found that students in Group A were more likely to include higher quality evidence in their written arguments. Often when they included higher quality evidence they also provided evidence conclusion and/or reasoning. Although the curriculum included rating the quality of the evidence, I suggest more discussions are needed about the differences between evidences and explaining how evidence differs and why some evidence was better than other evidence. For instance, in HIV lesson 7 students were asked to respond to the questions, “Which evidence is most useful for helping you decide between the models? Explain why.” For both groups, students chose the higher quality evidence and, in their written arguments, used this evidence to justify choosing the better-supported model.

The final suggestion focuses on developing criteria to promote argument quality (Chinn et al., 2014) and including epistemic criteria for evidence quality. Although in previous studies argument quality was related to how well students understood the content (e.g., Osborne et al., 2004), I did not find this to be the case. Even though Group A’s understanding of evidence and content was less than Group B, I found that Group A was more likely to include higher quality evidence, reasoning and rebuttals in their arguments. How students evaluated and used evidence led to differences in argument quality more so than their understanding of the content. This is why I suggest the argumentation learning progression should include aspects that focus on argument quality. For instance, the learning progressions should include complexity of arguments changing from students using lower quality evidence to higher quality evidence, and an increase in including evidence conclusions. Argument quality criteria need to be developed and added to the argumentation learning progression, especially the epistemic criteria for evidence quality. Students can learn how to apply evidentiary criteria when evaluating and constructing verbal arguments (Ryu & Sandoval, 2012). The inclusion of evidentiary criteria may be useful for moving the quality of written arguments from simple to complex.

Conclusion

The purpose of this study was to examine change in students' arguments across most of the school year and address changes that occurred in their written arguments. As part of a model-based inquiry curriculum, students should participate in the evaluation of competing models and multiple pieces of evidence, and have opportunities to identify the link between a model and a piece of evidence. Even though both groups had a similar understanding of content, models, and evidence, Groups A and B differed in how they used evidence, the quality of evidence they used, and the inclusion of reasoning and contradictory-evidence arguments in written arguments.

In general, arguments where students described evidence and/or stated an evidence conclusion, they were more likely to include a reason in their argument and provide a rebuttal. Although providing a rebuttal is a higher order task than reasoning, model choice, evidence conclusions and evidence usage, I did find that students in both groups provided rebuttals. However, students in Group A began to provide rebuttals earlier in the curriculum than students in Group B. One suggestion for the difference between groups was that students in Group A used higher quality evidence earlier in the curriculum than students in Group B.

I found that the supports embedded in the curriculum were valuable for writing arguments, especially the argumentation rubric and MEL matrix. The MEL matrix was somewhat helpful in supporting students as they included reasoning and rebuttals in their arguments (especially towards the end of the curriculum). The introduction of the argumentation rubric helped Group A's arguments continue to increase, except for the genetics assessment when the argumentation rubric was not used. For Group A, there was an increase in instances when students' provided rebuttals and reasoning after the argumentation rubric was introduced. The argumentation rubric provided students with a structure that is similar to the claims, evidence, and reasoning framework (McNeill et al., 2006; McNeill & Krajcik, 2008a, 2008b; Krajcik & McNeill, 2015), and helped support students in understanding what components were needed to be included in the argument. However, the other supports (MEL matrix, evidence ratings and class criteria for model goodness) were needed to help students address the quality of the argument.

Both Berland and McNeill (2010) and Osborne and colleagues (2016) argumentation learning progressions focus on the structure of the argument. They suggest that change in argument quality was due to students' understanding of scientific content, but I did not find this. I found that students' usage of evidence and their ability to evaluate the quality of evidence explained differences in argument quality. When students included higher quality evidence in their written arguments they were more likely to include reasoning and rebuttals. Both learning progressions included students using evidence to support a claim, but did not discuss the quality of the evidence or how including lower quality evidence in an argument was different than including higher quality evidence.

Limitations

There were some limitations of this study including: (a) sample size, (b) inability to apply statistical analysis to determine difference in argument quality due to sample size, (c) limited number of arguments analyzed (six time points over 22 weeks of instruction), and (d) limited number of arguments written after the introduction of the final support (argumentation rubric). The sample of this study was 20 students purposefully selected from a larger sample of students. The sample size was small because I wanted to only evaluate students who made gains from pretest to posttest and completed all of the writing prompts. I wanted to evaluate how students' arguments change over time, so I analyzed their arguments at different points and had to focus on a smaller sample. Due to the small sample size, I was not able to run any tests for statistical significance. The final two limitations occurred due to evaluating arguments at only six time points (when argument generation opportunities existed in the curriculum and assessments). After the argumentation rubric was introduced, most students used the rubric for the final two lessons, but not the final two assessments (genetics assessment and posttest). For future work, I would like to analyze a larger sample and one or two additional written arguments after the final support was introduced.

References

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765-793.

Berland, L.K., & Reiser, B.J. (2009). Making sense of argumentation and explanation. *Science Education, 93*(1), 26-55.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191-216.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082-1112.

Bingle, W. H., & Gaskell, P. J. (1994). Scientific literacy for decisionmaking and the social construction of scientific knowledge. *Science Education, 78*(2), 185-201.

Bybee, R.W. (1987). Science education and the science-technology-society (S-T-S) theme. *Science Education, 71*(5), 667-683.

Chinn, C. A., Duncan, R. G., Dianovsky, M., & Rinehart, R. (2013). Promoting conceptual change through inquiry. In *International handbook of research on conceptual change* (pp. 539-559).

Chinn, C. A., Duncan, R. G., Hung, L. C-C., & Rinehart, R. W. (2014, April). *The quality of arguments in problem-based and inquiry learning environments.* Paper presented at the Problem-Based Education SIG, American Educational Research Association, Chicago.

Chinn, C. A., Duschl, R. A., Duncan, R. G., Buckland, L. A., & Pluta, W. J. (2008, June). A microgenetic classroom study of learning to reason scientifically through modeling and argumentation. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 3* (pp. 14-15). International Society of the Learning Sciences.

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*, 425-453.

Chinn, C. A., & Rinehart, R. W. (2016). Epistemic cognition and philosophy: Developing a new framework for epistemic cognition. *Handbook of epistemic cognition*, 460-478.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom mathematical practices. *The Journal of the Learning Sciences, 10*(1&2), 113-163.

Creswell, J.W. (2007). *Qualitative inquiry & research design: Choosing among five approaches.* (2nd ed.) Thousand Oaks, CA: Sage.

Davis, E. A., & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. *The journal of the learning sciences*, *13*(3), 265-272.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, *84*(3), 287-312.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, *46*(6), 606-609.

Duncan, R. G., & Rivet, A. E. (2018). Learning progressions. In F. Fischer, C.E. Hmelo-Silver, S.R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences.* New York: Routledge.

Duschl, R. A., Schweingruber, & H. A., Shouse, A.W. (2007). *Taking science to school: Learning and teaching science grades K-8*. Washington, D.C: National Academies Press.

Edelson, D. C. (2001). Learning-for-use: a framework for integration content and process learning in the design of inquiry activities. *Journal of Research in Science Teaching*, *38*(3), 355-85.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry based learning through technology and curriculum design. *Journal of the learning sciences*, *8*(3-4), 391-450.

Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 335–354). Cambridge, UK: Cambridge University Press.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, *88*(6), 915-933.

Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, *45*(8), 922-939.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of science*, *71*(5), 742-752.

Herrenkohl, L. R., & Guerra, M. R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and instruction*, *16*(4), 431-473.

Hofer, B., & Pintrich, P. (1997). The Development of Epistemological Theories: Beliefs about Knowledge and Knowing and Their Relation to Learning. *Review of Educational Research,67*(1), 88-140.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The elementary school journal*, *94*(5), 483-497.

Krajcik, J., & McNeill, K. L. (2015). Designing and assessing scientific explanations tasks. In *Encyclopedia of Science Education* (pp. 285-291). Springer Netherlands.

Krajcik, J., McNeill, K.L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education. 92*(1), 1-32.

Lehrer, R., & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*, *21*(1), 39-48.

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in elementary grades. In T. A. Romberg, T. P. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 29–54). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Schauble, L. (2006). Cultivating Model-Based Reasoning in Science Education. In R. K. Sawyer (Ed.), *The Cambridge handbook of: The learning sciences* (pp. 371-387). New York, NY, US: Cambridge University Press.

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development, 23*(4), 512–529.

Lucas, D., Broderick, N., Lehrer, R., & Bohanan, R. (2005). Making the Grounds of Scientific Inquiry Visible in the Classroom. *Science Scope*, *29*(3), 39-42.

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*(2), 233-268.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, *48*(7), 793-823.

McNeill, K. L., & Krajcik, J. (2008a). Assessing middle school students' content knowledge and reasoning through written scientific explanations. *Assessing science learning: Perspectives from research and practice*, 101-116.

McNeill, K. L., & Krajcik, J. (2008b). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of research in science teaching*, *45*(1), 53-78.

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

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

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*(2), 93-127.

Metz, K. E. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction, 22*, 219-290.

Metz, K. E. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal, 109*, 138-161.

NGSS Lead States (2013). Next Generation Science Standards: For States, By States.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463-466.

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.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of research in science teaching*, *53*(6), 821-846.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction?. *Science Education*, *95*(4), 627-638.

Passmore, C. & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching*, *39*(3), 185-204.

Passmore, C. & Svoboda, J. (2012). Exploring opportunities for argumentation in modelling classrooms. *International Journal of Science Education, 34*(10), 1535-1554.

Patton, M. Q. (2002). *Qualitative research and methods*. John Wiley & Sons, Ltd.

Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, *48*(5), 486-511.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., ... & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, *13*(3), 337-386.

Rinehart, R. W., Duncan, R. G., & Chinn, C. A. (2014). A scaffolding suite to support evidence-based modeling and argumentation. *Science Scope*, *38*(4), 70.

Rinehart, R. W., Duncan, R. G., Chinn, C. A., Atkins, T. A., & DiBenedetti, J. (2016). Critical design decisions for successful model-based inquiry in science classrooms. *International Journal of Designs for Learning*, *7*(2).

Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, *96*(3), 488-526.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of research in science teaching*, *41*(5), 513-536.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*(4), 634-656.

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

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science education*, *88*(3), 345-372.

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist*, *49*(2), 139-152.

Schwarz, C. V., & Gwekwerere, Y. N. (2007). Using a guided inquiry and modeling instructional framework (EIMA) to support preservice K-8 science teaching. *Science education*, *91*(1), 158-186.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Archer, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-654.

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and instruction*, *23*(2), 165-205.

Shea, N. A., & Duncan, R. G. (2013). From theory to data: The process of refining learning progressions. *Journal of the Learning Sciences*, *22*(1), 7-32.

Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching*, *49*(2), 141-165.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610-631.

Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. *How students learn*, 515-565.

Tali, T., Krajcik, J. S., & Blumenfeld, P. C. (2006). Urban schools' teachers enacting project-based science. *Journal of Research in Science Teaching*, *43*(7), 722-745.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). “Mapping to know”: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, *86*(2), 264-286.

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

Appendix A

Focal Lessons

Unit 3: Cells, Lesson 1: Chloroplasts

The chloroplast lesson was discussed in detail in chapter 3. The driving question for this lesson was: “what do chloroplasts do?” Students were introduced to four pieces of evidence. The purpose of the first four pieces of evidence was to provide students with background content about chloroplasts to help them develop their initial model of chloroplast function. After the development of the model, students discussed their models in groups and were introduced to the final pieces of evidence. The last two pieces of evidence provided additional information about the function of chloroplasts. Students used the additional evidence to revise their initial models. Students were then asked to evaluate their model and individually answer the question: “explain why your model is a good model. Explain what evidence supports your model and explain in detail how the evidence supports your model”. This question was chosen for analysis because it is similar to the question students were asked in the pretest and posttest. Students need to evaluate a model and use evidence to support their reasoning.

Unit 3 Cells, Lesson 2 Nucleus

This second lesson in the Cells unit focuses on the role of the nucleus. Students answered the driving question: “what does the nucleus do?” Students began this lesson by generating criteria for rating evidence. Students were introduced to the statement, “When evidence is good, you can believe the evidence conclusion; and when it is bad you cannot”. Students learned to rate each piece of evidence using the numbers 0, 1 and 2 (from worst quality to best quality). A rating of 0 is considered very bad evidence and is evidence that should be ignored. A rating of 1 is considered not very good evidence, but not totally bad. Students were told that this is evidence that could be believed a little, but to be careful. A rating of 2 is considered good evidence. Students were told that the evidence described a correct conclusion. By rating the goodness of evidence, students have a better understanding of differences in the quality of each piece of evidence and identify the better quality evidence to use to support their model choice.

Students were then introduced to two models, control center model and instructions model. The models were designed to answer the questions, "what does the nucleus do? Use three pieces of evidence determine which model is better, and then revise the better model to make it even better". Students were next introduced to five pieces of evidence and asked to rate the quality of each piece of evidence individually (for evidence 1 and 2). Students also completed a MEL matrix linking evidence to the models. For lessons that occurred after this lesson, the evidence quality rating was included in the MEL matrix and sometimes for individual pieces of evidence. This lesson also encouraged students to discuss how a piece of evidence supports a model and explore the relationship between a model and the evidence.

With evidence 1, Dolly the sheep, students viewed a computer simulation in which scientists took the nucleus out of an egg cell from a sheep with a black face and replaced it with a nucleus from the egg cell of a sheep with a white face. The fertilized egg cell was transplanted into the womb of a female sheep. The evidence described that the baby sheep had a white face. With evidence 2, glowing cats, groups of students read a Scientific American blog. The blog described a study in which scientists took the GFP gene from a jellyfish and injected it into a cat's egg cell. The fertilized egg cell was transplanted into the womb of a cat and the resulting kitten produced GFP protein in its cells. The kitten glowed under UV light just like jellyfish. With evidence 3, clawed frogs, students read a report that some clawed frogs give birth to abnormal tadpoles that have many cells without a nucleus. As a comparison, students observed scientists' drawings of normal and abnormal tadpoles. Students observed that tadpoles without a nucleus die before becoming frogs. With evidence 4, website, students read information from the website www.Ineedhelp.com where someone asked, "What is the function of a nucleus?" The answer to this question, the website author replied that the nucleus is the control center (brain) of the cell and it contains DNA. The author continued that the cell's DNA directs the function of cell structures. Finally with evidence 5, diabetes, students read a study about diabetes. Scientists posed a hypothesis in the study that people with Type 1 diabetes have a mutated DRB gene and do not produce the DRB protein. In these cases, people would not produce enough insulin, resulting in diabetes. To test this, scientists in the study examined 10 healthy people and 10 people with Type 1 diabetes to see whether or not they had the gene mutation and if they produced the protein. Out of the 10 people with Type 1 diabetes, scientists found that eight of them had no DRB gene and produced no DRB protein. Out of the 10 healthy people, scientists found that nine of them did have the DRB gene and produced the DRB protein.

After students read and evaluated each piece of evidence, based on the quality of the evidence and relationship to the models, they evaluated the two models. Students were asked to discuss in pairs which model is better based on the evidence presented. They next individually revised the model chosen to make it an even better model. Students were then asked to explain their reasons for changes they made and answered the question: “explain why your model is a good model. Explain what evidence supports your model, and explain in detail how the evidence supports your model.” This was the same question asked in the chloroplasts lesson, but in this lesson students were asked to evaluate two models and chose the better model. After choosing the better model, students made revisions to the model to make it an even better model.

Unit 4: Genetics, Lesson 7: Existence of HIV Resistance

The driving question for this lesson was, “Does genetic resistance to HIV exist?” The lesson began by introducing two models, genetics resistance to HIV does not exist or genetics resistance to HIV does exist. Students were asked to select the best model. Students were then introduced to four pieces of evidence and asked to evaluate each piece of evidence. Students then completed the MEL matrix. Throughout this lesson, students were asked questions to determine their understanding of the evidence, to determine their capacity to rate the quality of the evidence and, after learning about evidence 1 and 2, to determine their capacity to evaluate the models and choose the best model.. With evidence 1, Feline Immunodeficiency Virus (FIV) video, students observed a video in which a scientists was worried that large wild cats (cheetahs, lions, pumas) could contract FIV because house cats could easily contract FIV. He analyzed blood samples and found that many large, wild cats had FIV in their blood but were not affected by the virus. The wild cats had a genetic mutation generating resistance to the disease. Whereas, house cats do not have this genetic mutation and are susceptible to FIV. With evidence 2, students read an article in which a journalist interviewed fifteen different nurses and doctors at a health clinic. From the interviews, students could conclude that humans are not resistant to HIV because HIV is transmissible in different ways. One thing to note was a statement by the nurse that explained that providing medicine to HIV infected pregnant women reduced transmission of HIV to the baby. Without the medicine, the baby usually contracted the disease from the mother. Evidence 1 and 2 provided new information to students that genetic resistance to HIV may exist.

With evidence 3, Simian Immunodeficiency Virus (SIV), students learned that monkeys can be infected with SIV and the virus is similar to HIV in humans. Some monkeys seem to be resistant to SIV even when exposed to the virus. Scientists conducted four breeding experiments with eight parent monkeys and found that if both parents are resistant to SIV then the offspring are resistant too. If either or both parents are susceptible to SIV, the offspring is also susceptible. Finally with evidence 4, Dr. Paxton's study, students learned that Dr. Paxton's team of scientists studied a group of 25 people who had been exposed to HIV many times. Despite exposure, the study participants were HIV negative. The white blood cells showed that they had some resistance to HIV. After students learned about and discussed each piece of evidence, they were asked to complete the MEL matrix and a model evaluation task. Students were asked to "write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer." Students were provided with an argumentation rubric while answering this question. This question was similar to the pretest and posttest question and provided students the opportunity to write an argument and use evidence to support their model choice.

Unit 4: Genetics, Lesson 8: Mechanism of HIV resistance

The driving question for this lesson was, "Why are some people resistant to HIV?" The lesson began by introducing two models, attack and destroy model and keep it out model. Students were then introduced to four pieces of evidence. As with lesson 7, students were introduced to pieces of evidence that range in quality and asked to evaluate each piece of evidence. Students then completed the MEL matrix. With evidence 1, the Burke family, students read a newspaper article about the Burke family. Everyone in the family was infected with HIV, except one of the children - Nikki. One member of the family believed that Nikki had a special gene that led to the creation of CCR5 protein that kills HIV cells. With evidence 2, comparison of white blood cells, students observed a computer simulation that compared the white blood cells evidence. A team of scientists studied the white blood cells of two people. Person 1 was genetically immune to HIV, and person was sick with HIV. They found the immune person's white blood cells do not have a protein called CCR5 and the HIV infected person's white blood cells had the CCR5 protein. The two pieces of evidence contradicted each other, so when students evaluate the quality of the evidence they also had to consider the source of the evidence.

With evidence 3, Hepatitis B and interferon, students read a health science news blog that explained how the body fights the Hepatitis virus. The blog explained that Hepatitis B is a disease caused by a virus that attacks the liver. Interferon is a protein produced by cells when bacteria or viruses attack them. Interferon acts by activating immune cells to destroy viruses. The scientists used Interferon to treat 12 patients with Hepatitis B. The patients got injections every two weeks for a year. After the treatment, three of the patients were cured of Hepatitis B, five Patients had a very low number of Hepatitis viruses, and four patients had a very high count still. Finally with evidence 4, Jen's favorite genes, students read a blog in which the author described a study of four different proteins: CD3, CCR5, CCR7 and CD8. The author tested a chemical mixture that she knew for sure had all four kinds of proteins as a check. She tested the cell membranes of 100 people who were infected with HIV and 10 who were not. She found that cell membranes from people resistant to HIV did not have the CCR5 protein while cell membranes from those not resistant had CCR5 protein. She also found that fluid leftover after taking out the cell membrane from both types had no CCR5 while the chemical mixture did. After students learned about and discussed each piece of evidence, they were asked to complete the MEL matrix and a model evaluation task. Students were asked to "write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer." Students were provided with an argumentation rubric while answering this question. This question was similar to the pretest and posttest question and provided students the opportunity to write an argument and use evidence to support their model choice.

Unit 5: Evolution, Lesson 9: Finches

The driving question for this lesson was, "What happened to the ground finches of the Galapagos Islands?" Students were presented with information that researchers studied a population of ground finches that live on the Galapagos Islands. A severe drought lasting several years caused a very large drop in the finch population. After the drought, the researchers discovered that the population of finches had changed. The purpose of this lesson was to encourage students to use the evidence to determine how the population changed and why the change occurred. Students were asked to work in pairs and evaluate evidence 1-5 on the computer. The five pieces of evidence provided background information about the island the finches inhabited, ground finches, predators that eat finches, seeds and differences in the finches before and after the drought. The order they studied the evidence was up to the pairs, and after studying the five pieces of evidence they were asked to individually develop an initial model. With evidence 1, what is their island like?, students learned about the conditions of the island, Daphne Major, where the finches lived. It is a small, warm island, gets a small amount rain from December to May and it is very rocky and dry. The island is not inhabited by people, and the only people who visit are scientists. With evidence 2, general information about group finches, students learned that finches eat a variety of seeds, and fly around the island but do not fly to other islands. They live around two years, and lay on average three eggs per month. Some genetic traits of the finches include beak size, wingspan, foot length, body color and beak color. With evidence 3, students learned about the predators on the island before and during the drought. A scientist counted how many of the two types of predators, hawks and short-eared owls, she could find that eat ground finches before and during drought. The number of predators before and during the drought was similar in number. With evidence 4, students learned about the seeds on the island before and during the drought. Plant scientists studied the small, soft seeds finches ate and found that during the drought there was an increase in the number of seeds with hard shells and a decrease in the number of seeds with soft shells. With evidence 5, differences in finches before and after the drought, students learned that scientists collected data about different traits of medium sized ground finches before the drought and after the drought. After the drought, the number of finches and their weight decreased, but beak size, wingspan and foot length stayed the same. After studying the five pieces of evidence, students were asked to individually develop an initial model that showed how the population of ground finches changed after the drought and why the population changed. Students were encouraged to use the general natural selection model developed to help them. After developing the models, students explained their models to each other. As part of the discussion, students were asked to address if their model fit with the class's criteria for good models and if there were changes they could make to their model to fit the criteria even better.

After the discussion on their initial models, students studied evidence 6-9 in pairs. With evidence 6, variation in finch beaks and how the beaks are used, students learned about a professor who took many pictures of finches before, during and after the drought and observed a range of different beak sizes. Finches with small and blunt beaks could crack open only small, soft seeds. Finches with bigger, more pointy beaks could crack open both small, soft seeds and larger, harder seeds. With evidence 7, observation of birds fighting, students learned about a couple that visited the island in 1978 during the drought. They saw two birds fighting over a seed and the larger bird drove the smaller bird away and ate the seed. With evidence 8, seeds from four plants before and during the drought, students learned about a research team that surveyed four seed bearing plants (cactus, chamae, portulaca, and tribulus) that ground finches ate before and during the drought. During the drought the number and type of seeds diminished for medium and small, soft seeds. No small, soft seeds were found during the drought and a small amount of medium seeds were found. The large, hard seeds were about the same in number before and during the drought. With evidence 9, weight of finches that survived and died during the first 6 months of the drought, students learned about a research team that weighed birds that survived the drought and birds that did not survive the drought. The research team found the average weight of living and dead birds was about the same. Birds that survived the drought, on average, were slightly heavier.

After students learned about and discussed the final four pieces of evidence, the teacher introduced two models that could also explain what happened to the population of ground finches because of the drought. Students were then asked to complete the MEL matrix and a model evaluation task. Students were asked to "write an argument to support your model. Write to someone who may not agree with you. Use your argumentation rubric to help you write your argument." This question was similar to the pretest and posttest question, and provided students the opportunity to write an argument and use evidence to support their model choice.

Appendix B

Lesson 7: "Does Genetic Resistance to HIV Exist?"

Before seeing any evidence, which model do you think is the right one? Circle your selection.

Model 1: Genetic resistance to HIV does not exist.	Model 2: Genetic resistance to HIV does exist.

1. Write your reasons for the model you picked.

____

Do Now

2. Explain the difference between HIV and AIDS.

Evidence 1 – FIV Video

<u>Video Summary</u>: The following is a summary of the video about FIV in cats.

<u>Introduction</u>: FIV stands for Feline Immunodeficiency Virus. FIV is a virus that attacks the immune system in house cats in a way that is similar to how HIV attacks the immune system in humans.

FIV was first observed in house cats, also called domestic cats. Dr. Stephen O' Brien noticed that house cats could get FIV very easily, and he was worried that FIV would spread from house cats to the large wild cats like cheetahs, lions, and pumas. Many of these species of large wild cats are endangered and could become extinct. Dr. O'Brien was afraid that many of these endangered species could die out if they were exposed to FIV.

<u>Method</u>: Dr. O'Brien gathered blood samples from thousands of large wild cats from around the world. He analyzed these samples. He used well known, reliable techniques for analyzing the blood for the presence of the virus.

<u>Results</u>: Most large wild cats like cheetahs, lions, and pumas already had FIV in their blood. However, they were not negatively affected by it because they possessed a genetic mutation that makes them resistant to the disease. Even though large wild cats get the virus, they do not become sick. Unlike wild cats, house cats do not have this genetic mutation and are not resistant to the disease. When house cats get infected with FIV, they often become very sick and can die.

<u>Conclusion</u>: From the blood samples of thousands of wild cats and house cats, Dr. O'Brien concluded that wild cats are genetically resistant to FIV, and house cats are not genetically resistant to FIV.

3A. Most wild cats who get FIV become sick and can die. True False

3B. House cats do not get the FIV resistant gene. True False

4. Geeta and Jose are arguing about this evidence. Circle the one you agree with the most.

A. Geeta thinks cats are mammals like humans and research on cats is useful for understanding HIV.

B. Jose thinks cats are different from humans and research on cats is not useful for understanding HIV.

C. I don't agree with either of them.

Explain your choice for your answer to question 4.

Evidence 2 – Greater Area Health Clinic

Interview Report:

It is common for people with HIV to be treated in health clinics. A journalist interested in whether some people are genetically resistant to HIV interviewed the nurses and doctors at the Greater Area Health Clinic.

The journalist interviewed fifteen different nurses and doctors at this health clinic. Here are a few things the interviewees said:

Dr. Gutierrez: "It used to be, back in the 1980s, people would come in with HIV and there was very little that we could do to help. In the 1990s we developed medicine that attacked HIV in the blood stream. This reduced the infection but it didn't cure it. People taking the medicine people live longer than people who don't take the medicine."

Nurse Singh: "I have worked in the labor and delivery ward for twenty-seven years. It used to be that if a pregnant woman came in and she had HIV, the baby would usually get the disease too. Now we can give mothers some medicine that reduces the chance the baby will get it. If the mothers don't get the medicine, the babies will still usually get the disease."

Dr. Morse: "With my patients I try to stress the point that everyone can get HIV. You can get it from injecting drugs with contaminated needles or having sex with someone who has the disease."

Lab Assistant Feld: "I have worked in the blood lab for about five years. We check patients' blood for HIV. The test is about 99% accurate. I have never met anyone who is resistant to HIV. We have had some patients who thought they were resistant because they injected drugs for a long time and didn't get it. But within a few years they eventually got HIV."

5. How do you rate the quality of this piece of evidence (0, 1, or 2)? ☐

Give reasons for your rating.

Arrows Diagram

Evidence Goodness Rating	Model 1: Genetic resistance to HIV does <u>not</u> exist.	Model 2: Genetic resistance to HIV does exist.
 **1. FIV Video**		
2. Greater Area Health Clinic: <u>Interview Report</u>		

6. For evidence 1 and evidence 2 rate the evidence (0, 1, or 2) and then draw an arrow for how the evidence relates to each model. Make sure to put ratings and arrows for **<u>both</u>** pieces of evidence.

Support	$\longrightarrow$
Strongly Support	$\Longrightarrow$
Contradict	$\not\longrightarrow$
Strongly Contradict	$\not\Longrightarrow$
Irrelevant	$\dashrightarrow$

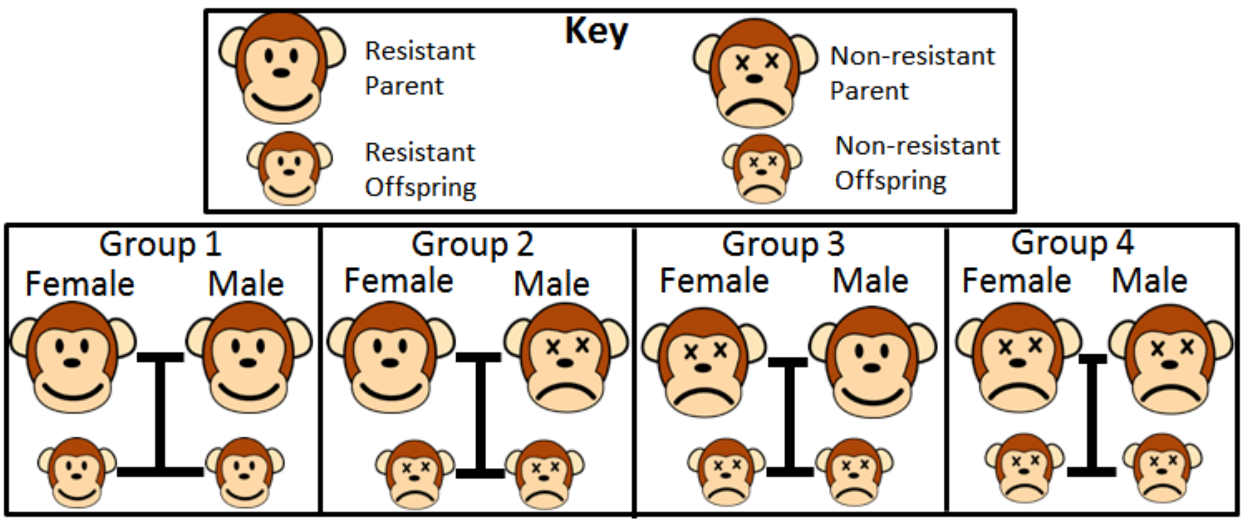
7. How does evidence 2 relate to Model 2? Give reasons for the arrow you chose.

8. Which model is better? Circle your selection.

Model 1: Genetic resistance to HIV does <u>not</u> exist.	Model 2: Genetic resistance to HIV does exist.

9. Which model do you think is best and why? Be sure to give reasons for your answer.

Evidence 3 – SIV
<u>Introduction</u>: Monkeys can be infected by SIV (Simian Immunodeficiency Virus). SIV is similar to HIV, the virus found in humans. Some monkeys seem to be resistant to SIV even when exposed to the virus. Resistant monkeys have SIV in their blood, but they do not develop AIDS. Monkeys that are not resistant to SIV develop AIDS and get sick.
<u>Method and Results</u>: Scientists did four breeding experiments with eight parent monkeys. The groups were completely separated so that they did not have contact with monkeys outside of their group. All monkeys were tested for SIV resistance using high-quality blood tests.
Group 1: A resistant mother and resistant father have resistant offspring
Group 2: A resistant mother and non-resistant father have non-resistant offspring
Group 3: A non-resistant mother and resistant father have non-resistant offspring
Group 4: A non-resistant mother and non-resistant father have non-resistant offspring



10a. Is SIV resistance in monkeys genetic? Circle your answer.

A. No it is not genetic.
B. Yes it is genetic and resistance is a dominant trait.
C. Yes it is genetic and resistance is a recessive trait.

10b. Explain why it is or is not genetic based on the results of this study. Give reasons for your answer.

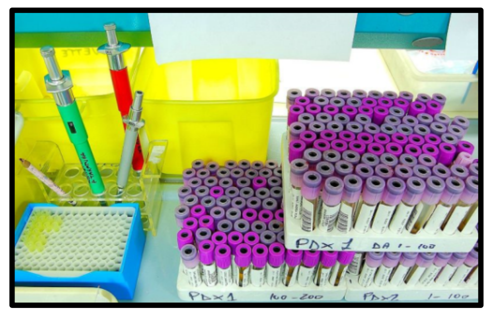
Evidence 4 – Dr. Paxton's Study

<u>Introduction</u>: During the 1990s Dr. Paxton heard that there were some people who had been exposed to HIV, but didn't develop AIDS. He wanted to see if their immune system cells would be resistant to HIV if they were exposed to it again. People who have unprotected sex or inject illegal drugs are more likely to get HIV, so they decided to study these people.

<u>Method</u>: Dr. Paxton and his team of researchers studied a group of 25 people who had been exposed to HIV many times. Despite many exposures, the people in the study were HIV negative, which means that there was no HIV in their blood.

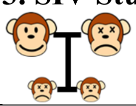
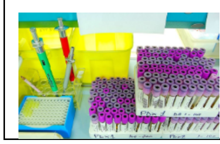
The researchers used white blood cells taken from these 25 people. The white blood cells were exposed to different levels of HIV in a test tube.

<u>Results</u>: All 25 peoples' white blood cells showed some resistance. Some people had immune system cells that were resistant to very high levels of HIV in the test tube.



11. What conclusion do you draw from this study? Explain your answer.

Arrows Diagram

Evidence Goodness Rating	Model 1 Genetic resistance to HIV does <u>not</u> exist.	Model 2 Genetic resistance to HIV in does exist.
1. FIV Video 		
2. Greater Area Health Clinic: <u>Interview Report</u>		
3. SIV Study 		
4. Paxton Study 		

12. For **<u>all</u>** the pieces of evidence make sure to rate them (0, 1, or 2) and draw an arrow for how the evidence relates to each model.

Support	→
Strongly Support	⇒
Contradict	↛
Strongly Contradict	⇏
Irrelevant	⇢

13. Which model is better? Circle your selection.

Model 1: Genetic resistance to HIV does <u>not</u> exist.

Model 2: Genetic resistance to HIV does exist.

14. Which evidence is most useful for helping you decide between the models? Explain why.

15. Write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer.

16. In pairs discuss the following four questions.

    Which evidence is the best evidence?
    Which evidence helps you the most to decide which model is better?
    What is the best argument for Model 1?
    What is the best argument for Model 2?

17. As a class discuss which model is best. Give lots of reasons for your answers.

Appendix C

Pieces of Evidence for Each Writing Prompt

Assessments

Assessment	Writing Prompt	Pieces of Evidence (E)
Pretest/Post test Falling Leaves	Which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer.	E1: Leaves begin falling from trees in September. By late November, almost all the leaves have fallen, and the trees are bare. E2: Leaves are made up of different types of cells. When leaf cells are damages, the leaf may die. E3: Scientists compared leaves from trees in summer and autumn. Nearly all the cells of leaves in spring were alive, but there were more dead cells in leaves that had just fallen. E4: Photographs from a powerful microscope of the leaf stem during the summer and during the autumn. (there were differences in the leaf stem between summer and autumn) E5: Scientists found that most leaves fell before the temperature was below freezing.
Pretest/Post test Muscle Pain	Which do you think is the better model for the problem? Write at least three (3) detailed reasons for your answer.	E1: After heavy exercise, the soreness does not occur right away. Usually the body is core from 24 to 72 hours after heavy exercise. E2: Nerve cells are attached to muscle cells. They can sense pressure and other changes around them. E3: People who fell pain after two days of heavy exercise had swollen muscles. People without pain did not have swollen muscles. E4: Scientist measured how much lactic acid there was in people’s cells before and after they exercised. The average amount of lactic acid in cells was similar before and 2 hours after exercise. E5: Scientist looked carefully at muscle fibers under a powerful microscope, and after hard exercise, muscle fibers are no longer straight.
Genetics Assessment	Which do you think is the best explanation of DEB, the separatin protein	E1: DEB runs in families. DEB may skip a generation. However, if both parents have DEB all their children will also have it.

	explanation or the connectin protein explanation? Write at least three (3) detailed reasons for your answer.	E2: Scientists compared skin samples from normal and affected individuals. They found large gaps between the epidermis and dermis layers in the samples from people with DEB. The samples from normal people did not have such gaps. E3: The skin has several layers. E4: Studies of the genetic material of normal and affected people show that they both have the same amount of DNA. E5: Scientists compared the different types of proteins found in the skins of normal and affected people. Affected people were missing one type of protein. E6: Scientists injected ten DEB patients with connectin protein. They found that 80% of the patients got better and their skin did not produce blisters when bruised.

Lessons

Unit, Lesson	Writing Prompt	Pieces of Evidence (E)
Cells, Chloroplasts	Individually, explain why your model is a good model. Explain what evidence supports your model, and explain in detail how the evidence supports your model.	E1: Elodea Experiment, chloroplasts are found in plants cells and are green. E2: Bacteria Experiment: Engelmann found that the bacteria went to the places around the edge of the cells next to the chloroplasts. E3: Animal cells had no chloroplasts. Cactus root and oak tree trunk cells had few chloroplasts. Oak tree leaves, moss, tulips, cactus and algae cells had many chloroplasts. E4: The green parts of the leaf turned dark blue and contained starch. The white parts of the leaf turned brown and contained no starch. The covered-up parts of the leaf turned brown and contained no starch.) E5: Barley plants with normal chloroplasts make large amounts of glucose. Barley plants with mutated chloroplasts make small amounts of glucose. Barley plants with very mutated chloroplasts make no glucose.) E6: In the light, the amount of oxygen in the jar got bigger, and the amount of carbon dioxide got smaller. In the dark, the amounts of oxygen and carbon dioxide did not change.
Cells, Nucleus	23. Individually, explain why your model is a good	E1: The baby sheep had a white face.

	model. Explain that evidence supports the model, and explain in detail how the evidence supports your model.	E2: The GFP gene injected into the mother cat's egg caused the baby kitten to be born with the GFP gene. GFP protein was produced in the kitten's cells. The kitten glowed under UV light just like jellyfish do. E3: Findings: Scientists made drawings showing how different the normal and abnormal tadpoles look from each other as they grow older. Also, tadpoles without a nucleus die before becoming frogs. E4: Findings: The answer to this question said that the nucleus is the control center (brain) of the cell and it contains DNA. The cell structures do as they are told by the DNA. E5: Findings: Out of the 10 people with Type 1 diabetes, scientists found that 8 of them had not DRB gene and produced no DRB protein. Out of the 10 healthy people, scientists found that 9 of them did have the DRB gene and produced the DRB protein.
Genetics, Lesson 7 (Does genetics resistance to HIV exist?)	15. Write an argument to support your model. Write to someone who may not agree with you. Give detailed reasons for your answer.	E1: (FIV video): FIV, Feline Immunodeficiency Virus, is a virus that attacks the immune system in house cats in a way that is similar to how HIV attacks After analyzing, Dr. O'Brien concluded that wild cats are genetically resistant to FIV, and house cats are not genetically resistant to FIV. E2: (Greater Area Health Clinic): Humans are not resistant to HIV, because you can get HIV in multiple ways. E3: (SIV study): Monkeys can be infected by SIV (Simian Immunodeficiency Virus). SIV is similar to HIV, the virus found in humans. The only resistant offspring came from a pair of two resistant parents. E4: (Dr. Paxton's Study): Dr. Paxton and his team of researchers studied a group of 25 people who had been exposed to HIV many times. Despite many exposures, the people in the study were HIV negative. All 25 people white blood cells showed some resistance, with some being resistant to very high levels of HIV.
Genetics, Lesson 8 (Mechanism of HIV resistance)	16. Write an argument to support the model you chose. Write to someone who may not agree with you.	E1: (Burke family): The Burke are a family of 4. Everyone in the family except one of the children, Nikki, has been infected by HIV. One person believes she has a gene that leads to the creation of CCR5 protein, which goes on to kill HIV cells. E2: (Dr. Aller's study on white blood cells): In December of 2010 Dr. Kristina Allers and her team of scientists carefully studied the white blood cells of two people. They found that (immune) Person #1's white blood cells do not have a protein called CCR5 and (not immune) Person #2's white blood cells do.

		E3: (Hepatitis B and interferon): Scientists treated 12 patients who were infected with Hepatitis B with interferon. They got injections every two weeks for a year. After the treatment, 3 people had no Hepatitis B, 5 people had a very low number, and 4 had a very high count still. E4: (Jen's favorite genes): She found that membranes from people resistant to HIV did not have the CCR5 protein while membranes from those not resistant did. She also found that fluid leftover after taking out the membrane from both types had no CCR5 while the chemical mixture did.
Evolution, Lesson 9 (Finches)	10. Pick which model you think is better, the Beak Size Model or the Fighting Model: Write an argument to support the model you chose. Write to someone who may not agree with you. Use your argumentation rubric to help you write your argument.	E1: Information about Daphne Major and what the island is like. E2: General information about ground finches. E3: Predators on the island before and during the drought. E4: The seeds on the island before and during the drought. E5: Differences in finches before and after the drought. E6: Variation in finch beaks and how the beaks are used. E7: Observation of birds fighting. E8: Seeds from four plants before and during the drought. E9: Weight of finches that survived and died during the first 6 months of the drought.

Appendix D

Content, Model, and Evidence Understanding Questions

Content Understanding Questions

Nucleus Lesson:

1. What is the function of the nucleus?

2. What would happen if you took the nucleus of a mouse egg and put it in a chicken egg?

3. Some muscle cells have more than one nucleus. How would muscle cells with more than one nucleus be different than cells that have only one nucleus?

Evolution Lesson:

1. How did the environment change on the island?
   a. the number of predators that ate finches got larger
   b. the size of the finches' beaks got larger
   c. the number of seeds with soft shells got smaller
   d. the amount of rainfall got larger

2. How did the population of finches change after the environmental change?
   a. after the change, there was a larger percentage of finches with big beaks
   b. after the change, all the finches had big beaks
   c. after the change, some of the finches grew bigger beaks
   d. after the change, all of the finches grew bigger beaks

3. Why did the population of finches change?
   a. the birds with bigger beaks were more likely to die before having offspring
   b. the birds with smaller beaks made their beaks larger so they could survive to have more offspring
   c. the birds with bigger beaks were more likely to survive and have more offspring with bigger beaks
   d. the birds with smaller beaks wanted their offspring to survive, so they had offspring with larger beaks

4. In this population of finches:
   a. all the finches have beaks that are the same size
   b. there is variation in the beak size of finches, but in other ways, the finches are all the same
   c. the finches vary in many different ways, including beak size

5. What type of variation in finches was passed to the offspring during the drought?

   a. offspring inherited all the genetic traits of the parent finches that survived the drought b. offspring inherited only the genetic traits they needed to survive in the drought
   c. offspring inherited the genetic traits of all the finches that were alive at the beginning of the drought

6. During this population change:

   a. some birds grew larger beaks
   b. some birds grew smaller beaks
   c. no birds grew larger or smaller beaks

Posttest:

1. What is the function of chloroplasts in cells? Explain in as much detail as you can.

2. Explain why a plant in a room full of light would live longer than plant in a dark room. Use what you have learned about chloroplasts to answer this question.

3. What is the function of the nucleus in cells? Write in as much detail as you can.

4. Jellyfish can glow in the dark. If scientists take the gene that makes the jellyfish glow and put it into cells of a piglet, what will happen to the piglet?

   a. The jellyfish gene tells the piglet cells to glow in and the cells make the piglet glow.
   b. The piglet cells use the jellyfish gene to make a protein that makes the piglet glow.
   c. The jellyfish gene doesn't work in a piglet and therefore the piglet doesn't glow.
   d. The is not possible because jellyfish do not have genes.

5. Which organelle do animal and plant cells not have in common?

   a. nucleus
   b. chloroplasts
   c. mitochondria
   d. animal and plant cells contain the same organelles

6. Which of the following statements is correct in explaining the size of the nucleus compared to the cell?

   a. the cell is smaller than the nucleus
   b. the cell is larger than the nucleus
   c. the cell and the nucleus are the same size

7. Genes give instructions to make different chemical structures called?

   a. nuclei
   b. proteins
   c. vacuoles
   d. DNA

8. In which kind of cell would you find the most amount of chloroplasts?

   a. cat skin cell
   b. human liver cell
   c. tree bark cell
   d. dandelion leaf cell

9. Which of the following is a correct statement about the function of chloroplasts?

   a. chloroplasts produce carbon dioxide
   b. chloroplasts produce proteins
   c. chloroplasts produce glucose
   d. chloroplasts produce light

10. Which of the following represents what happens in chloroplasts?

    a. oxygen and water and light energy are used to produce glucose and carbon dioxide
    b. hydrogen and oxygen are used to produce water
    c. carbon dioxide and water and light energy are used to produce glucose and oxygen
    d. oxygen and glucose are used to produce water and carbon dioxide and energy

Model Understanding Questions

Pretest:

1. Which explanation do you think is better? Circle your selection.

   a. poisonous chemicals explanation
   b. ice crystals explanation
   c. I can't tell right now

HIV, Lesson 7:

1. Which model is better? Circle your selection.

   a. Model 1: Genetic resistance to HIV does not exist.
   b. Model 2: Genetic resistance to HIV does exist.

HIV, Lesson 8:

1. According to the Attack-and-destroy Model, why are some people resistant to HIV?

   a. their genes give instructions to make extra white blood cells to attack HIV
   b. their genes give instructions to make a protein that turns on immune cells to attack HIV
   c. their genes do not give instructions to make a receptor protein that is used by HIV
   d. their genes give instructions to make a receptor protein that keeps HIV out

2. According to the Keep-it-out Model, why are some people resistant to HIV?

   a. their genes give instructions to make extra white blood cells to attack HIV
   b. their genes give instructions to make a protein that turns on immune cells to attack HIV
   c. their genes do not give instructions to make a receptor protein that is used by HIV
   d. their genes give instructions to make a receptor protein that keeps HIV out

3. Individually, answer this question: What is the best model of genetic HIV resistance? Circle your answer.

   a. Model 1: Attack-and-destroy
   b. Model 2: Keep-it-out

Genetics Assessment:

1. Which explanation do you think is better? Circle one.

   a. Separatin
   b. Connectin
   c. I can't tell right now

Evolution Lesson:

1. Pick which model you think is better, the Beak size model or the Fighting model?

   a. Beak size model
   b. Fighting model

Posttest:

1. Which model do you think is better? Circle your selection.

    a. poisonous chemicals model
    b. ice crystals model
    c. I can't tell right now

Evidence Understanding Questions

Chloroplasts Lesson:

1. Individually, answer this question. What do you conclude from this experiment?

2. Which of the following statements about the research report is correct?

    a. the plant cell the scientists is called Pseudomonas
    b. the bacteria try to get away from oxygen
    c. the bacteria go to places where there is a lot of oxygen
    d. the scientist used a magnifying glass to make his observations

3. What do you conclude from this experiment?

    a. shining light on cells cause oxygen to be used up near the chloroplasts
    b. shining light on cells causes oxygen to be produced near the chloroplasts
    c. shining light on cells does not affect oxygen around chloroplasts
    d. the bacteria used in this study need oxygen in order to stay alive

4. What do you conclude from this experiment?

5. Circle each part of that plant where the scientists tested for starch in this experiment. You will want to circle more than one of these plant parts.

    - green part of the leaf
    - white part of the leaf
    - part of the leaf covered by the paper
    - part of the leaf not covered by the paper
    - the stem

Nucleus Lesson:

1. Individually, answer this question: How good or bad is Evidence #1? Write your reasons for your answer. Write to someone who might disagree with you.

2. Individually, write your answer: which model (or models) does this evidence support, and why?

3. According to scientists, what makes the jellyfish glow?

   a. The GFP genes glow
   b. Their GFP protein glows
   c. Their GFP mitochondria glow

4. Which of these did the scientists do?

   a. They injected a glow-in-the-dark chemical into adult jellyfish
   b. They injected a glow-in-the-dark chemical into adult cats
   c. The put a cat gene into a jellyfish egg
   d. They put a jellyfish egg into a cat egg

5. According to the scientists, why did the cats glow?

   a. The GFP gene glows in cats
   b. The GFP gene gives instructions to make the GFP protein, and the GFP protein glows in cats
   c. The GFP gene does not do anything at all in cats, and this is why cats glow

6. What is the best conclusion for this study?

   a. The GFP gene gives instructions to make the GFP protein
   b. The GFP gene gives instructions to make the GFP protein, and the GFP protein makes animals glow
   c. The GFP gene does not make any proteins. No one knows why the cats and jellyfish glow

7. People get Type 1 diabetes because

   a. their bodies produce too much insulin
   b. their bodies do not produce insulin

8. According to the scientists' hypothesis, healthy people

   a. have a normal DRB gene which gives instructions to make the DRB protein
   b. have a normal DRB gene but no DRB protein
   c. have a mutated DRB gene and no DRB protein
   d. have a mutated DRB gene which gives instructions to make a lot of DRB protein

9. According to the scientists' hypothesis, people with Type 1 diabetes

   a. have a normal DRB gene which gives instructions to make the DRB protein
   b. have a normal DRB gene but no DRB protein
   c. have a mutated DRB gene and no DRB protein
   d. have a mutated DRB gene which gives instructions to make a lot of DRB protein

10. What is the best conclusion from this study about healthy people?

    a. most of them have normal DRB genes and DRB protein in their cells
    b. most of them have mutated DRB genes and no DRB protein in their cells

11. What is the best conclusion from this study about people with Type 1 diabetes?

    a. most of them have normal DRB genes and DRB protein in their cells
    b. most of them have mutated DRB genes and no DRB protein in their cells

HIV, Lesson 7:

1. Most of the wild cats who get FIV become sick and can die (T or F)

2. House cats do not get the FIV resistant gene (T or F)

3. Is SIV resistance in monkeys genetic? Circle your answer

   a. no it is not genetic
   b. yes it is genetic and resistance is a dominant trait
   c. yes it is genetic and resistance is a recessive trait

4. What conclusion do you draw from this study? Explain your answer.

5. Which evidence is most useful for helping you decide between the models? Explain why.

HIV, Lesson 8:

1. Why does Mr. Burke think that Nikki has stayed healthy?

   a. Nikki has not been exposed to the HIV virus
   b. Nikki has a protein that destroys the HIV virus
   c. Nikki has just been very lucky all her life
   d. Nikki is missing an important protein called the CCR5 protein

2. Individually, answer this question: How good or bad is Evidence #1? Write your reasons for your answer. Write to someone who might disagree with you.

3. What did the Dr. Kristina Allers test for?

   a. Whether or not the patients have white blood cells.
   b. Whether or not the patients have CCR5 protein
   c. Whether or not the patients have HIV

4. What is the best conclusion for this evidence?

   a. People who are infected with HIV do not have CCR5 receptors on their white blood cells
   b. People who are HIV resistant do not have CCR5 protein on their white blood cells
   c. People who are HIV resistant have CCR5 receptors on their white blood cells

5. Individually, answer this question: How is Evidence #2 related to Model 1 (supports, contradicts, etc.). Give reasons for your answer. Write to someone who might disagree with you.

6. Which is true of interferon?

   a. It is a protein that attacks and destroys viruses
   b. It is a kind of cell that attacks and destroys viruses
   c. It is a protein that turns on the immune cells which attack the virus

7. What did the scientists do in the study to try to help Hepatitis patients?

   a. They injected immune cells into 12 Hepatitis B patients for 1 year
   b. They injected interferon into 12 Hepatitis B patients for 1 year
   c. They did blood tests on 12 Hepatitis B patients for 1 year.

8. Individually, answer this question: How does this evidence relate to Model 1 (supports, contradicts, etc.). Give reasons for your answer. Write to someone who might disagree with you.

9. What conclusions do you draw from this study? Explain your answer.

Evolution Lesson:

1. How many mistakes are there in the following statement? The island called Daphne Major is very wet, contains a lot of grass, has no trees, has large plants, and has only wet seasons.

   a. 0
   b. 1
   c. 2
   d. 3

2. Which is true of the ground finches that live on Daphne Major?

   a. they eat insects, typically live three years, and fly around to several islands
   b. they eat seeds, typically live two years, and fly around to several islands
   c. they eat seeds, typically live three years, and fly only around Daphne Major
   d. they eat seeds, typically live two years, and fly only around Daphne Major

3. Which is true about the results of the study?

   a. the number of hawks seen decreased from 1975 to 1980
   b. the number of owls seen doubled from 1975 to 1980
   c. the number of hawks and owls stayed about the same from 1978 to 1980
   d. there were no owls or hawks seen in 1980

4. What is true about the results for the study?

   a. after the drought the abundance of soft seed shells increased
   b. after the drought the abundance of soft seed shells remained about the same
   c. after the drought, the average hardness of the seeds was higher
   d. after the drought, the average hardness of the seeds was about the same

5. What should the scientists conclude from the study?

   a. individual finches that survived the drought grew bigger beaks
   b. individual finches that survived the drought got bigger
   c. new generations of finches living after the drought had bigger beaks, on average
   d. new generations of finches living after the drought were bigger, on average