4-2016

Assessing the Impact of Historical Story Telling on Student Learning of Natural Selection

Janice Marie Fulford
Western Michigan University, fulford@kvcc.edu

ASSESSING THE IMPACT OF HISTORICAL STORY TELLING ON STUDENT LEARNING OF NATURAL SELECTION

by

Janice Marie Fulford

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Mallinson Institution for Science Education
Western Michigan University
April 2016

Doctoral Committee:

David W. Rudge, Ph.D., Chair
Marcia K. Fetters, Ph.D.
Patricia L. Reeves, Ed.D.

ASSESSING THE IMPACT OF HISTORICAL STORY TELLING ON STUDENT LEARNING OF NATURAL SELECTION

Janice Marie Fulford, Ph.D.

Western Michigan University, 2016

Research suggests that because of its historical nature, the learning of evolutionary biology is problematic compared to that of other science disciplines. While explanations used in historical sciences often employ historical narratives, which are distinct from narratives in other contexts, such as stories, the two types of narratives have structural similarities that suggest the potential role of stories based in the history of science for the teaching of evolutionary biology. Stephen Klassen, a prominent science educator, has studied how stories from the history of physics can promote the learning of and attitudes towards science. Klassen's pioneering work identifies structural components of stories (narrative elements) that give them explanatory power. To test Klassen's approach empirically, the present study employed an intervention (The Mystery Phenomenon (MP)) with reference to the history of research on industrial melanism (IM). The episode was chosen for study because it incorporates past scientists' theories and investigations on IM as a strategy to mitigate misconceptions. The efficacy of the unit was studied by means of a mixed-method approach that compared the learning outcomes and experiences of participants using two versions of the MP (one that employs a story that incorporates Klassen's structural components and another that did not). To

determine if the story approach impacted the learning of science content goals, participants in both groups took the Concept Inventory of Natural Selection (CINS) as a pre and post-test. A subset of participants also took part in semi-structured interviews to further clarify the analysis of the CINS results and also to assess the impact of Klassen’s structural components and student attitudes.

The study’s results demonstrates that the story version of the MP lesson yielded significant learning gains, and that some of the misconceptions explicitly discussed in the MP lesson displayed significant decreases. In addition, participants expressed positive attitudes to this lesson’s format as a mystery in reference to it as a teaching strategy. Finally, by employing two versions of the MP lesson, this study provides a systemic way for empirically testing the efficacy of stories.

ACKNOWLEDGEMENTS

I would like to thank my mentor and advisor, Dr. David W. Rudge for his dedication to me as a student and scholar. He truly understands the role of an advisor, and in that role takes a well-balanced approach. All feedback is given in the best interest of both myself and my research. I have thoroughly enjoyed and profited from my relationship with Dr. Rudge. I would also like to thank the other members of my committee, Dr. Marcia K. Fetters and Dr. Patricia L. Reeves. They have been an integral part of this project, reviewing my work and sharing their different strengths.

I am fortunate to be a part of an academic community that provides me with support in a variety of areas. My research group, headed by Dr. Rudge has given me emotional support, acted as a sounding board, and given invaluable advice on many occasions. The other students, staff and faculty at Mallinson Institute have given me a comfortable and safe space to learn, grow and pursue my research. I would like to thank my coders, Cody Williams and Tami Port; and Lisa Cronkhite-Marks for writing support. It takes a great deal of support to pursue a doctorate degree. I would to thank my parents, children and extended family and friends for their years of interest and support. No one has been a bigger cheerleader in my life than my husband, Glenn Fulford. He has been with me every step of the way and has never once complained about the hours I have devoted to this project.

Janice Marie Fulford

TABLE OF CONTENTS

Table of Contents-Continued

Table of Contents-Continued

CHAPTER

Table of Contents-Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The central role biology plays in modern life is reflected in national and state science education standards, and they explicitly include evolution (AAAS, 1993; NGSS Leads States, 2013). This is understandable since evolution ties all the diverse biological components together into an organized whole (Bishop & Anderson, 1990; Demastes, Settlage, & Good, 1995; Dobzhansky, 1973; van Dijk & Kattmann, 2009). In spite of the importance of this topic; students find it difficult to learn. These difficulties are compounded by students' alternative conceptions and their personal philosophies (Alters & Nelson, 2002; Bishop & Anderson, 1990). Indeed, Nehm and Reilly (2007) find that college biology students display limited knowledge of, and alternative conceptions about natural selection. These findings are not unique, Smith (2010) reports that most intervention studies fail to demonstrate acceptable levels of understanding in the majority of students, and Gregory (2009) summarizes over three decades of research in student learning of natural selection, and concludes that while several interventions have had some success in mitigating alternative conceptions, no complete solution has emerged. So in spite of sustained efforts, students are generally not learning evolution.

While there are many factors that contribute to difficulty in learning evolutionary biology, this study focuses on two factors that in combination create barriers for learning. The first factor is the historical nature of evolutionary biology which means scientists must rely on the past to find explanations for phenomena (Cleland, 2002). The second factor are learners' cognitive patterns that make evolutionary concepts difficult to

understand and accept (Evans, 2008). Evolutionary biology explanations have characteristics that are in direct opposition to learners' everyday thinking, which include misconceptions, and therefore can create resistance on the part of the learner. Historical sciences and cognitive biases are discussed next, followed by discussion on using stories as a strategy to help mitigate the issues caused by these factors.

Historical Sciences

The discipline of evolutionary biology is an historical science. There are some science disciplines, like geology, astronomy, paleontology and evolutionary biology that must look to the past for answers. This means that the type of evidential reasoning used in historical research is different than that in the physical sciences that employ classic experimental research (Cleland, 2002). These two different types of reasoning often overlap in a research program, however one type predominates in the historical disciplines and the other in the experimentally-based disciplines. Figure 1 shows the structure of historical explanations.

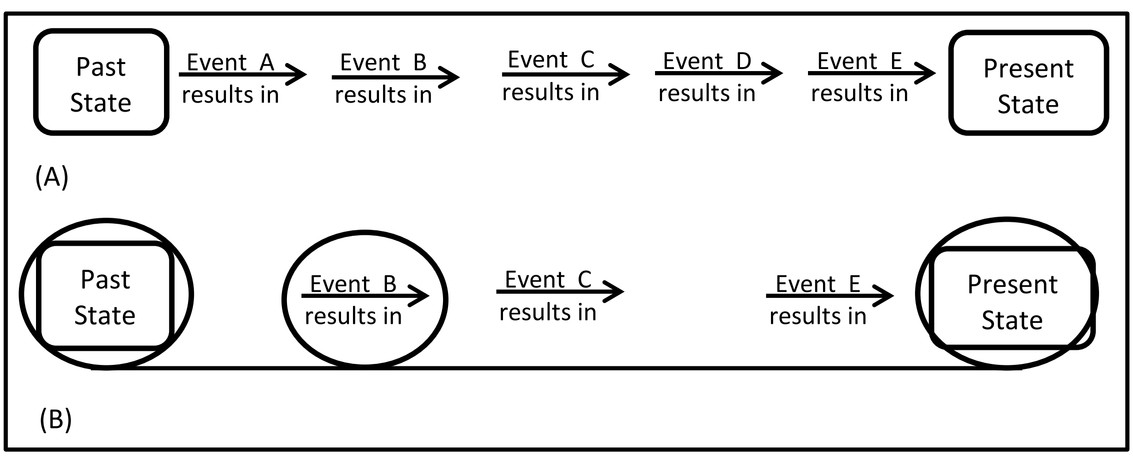


Figure 1. A causal chain for any present state needs a minimal amount of events.

Experimentally-based disciplines formulate hypothesis, test the phenomenon in a controlled setting, and use the results to make predictions. Experiments look for “regularities among event-types” (Cleland, 2002, p. 476). An event-type is recurring and not unique, therefore it can be repeatedly tested and verified, and results can direct future research and ultimately predictions can be made about these event-types. Cleland explains that researchers infer test conditions from their hypotheses and make predictions about what should occur if the hypotheses are correct. Experimental scientists pursue research programs that are a series of individual experiments, and there is interdependency between each experiment.

In the historical sciences, researchers observe current conditions or phenomena, formulate hypotheses, and look for evidence of past events to create a causal chain that explains the conditions or phenomena. Cleland (2002) explains that historical researchers are interested in event-tokens instead of event-types. Event-tokens are unique events or conditions that lead to a new events or conditions. The explanations of event-tokens often have a narrative quality because the causal chains linking past events to current situations are complex and the regularities get obscured by contingencies. Cleland states that the goal of the historical scientist is to discover a “smoking gun” that trace evidence that points to one hypothesis that is the most plausible explanation for all the traces. Historical sciences focus on explanations that provide unity to a diverse set of traces, and give them a causal framework. She reminds us that the “smoking gun” is a trace that identifies one causal chain as superior to competing explanations, not that it eliminates them, there is always the possibility that new evidence will make it necessary to rethink a current explanation. Cleland’s “smoking gun” can be related to Figure 1 in which “Event B’

represents the event-token in this specific chain of events that is the causal link between the past state and the present state.

Cleland (2002) concludes that experimental science and historical sciences have different methodologies to generate hypotheses because the former is moving from the present to the future, and the later is moving from the present to the past. Jeffares (2008) discusses Cleland’s work and agrees with her analysis that historical sciences use evidentiary traces to build a causal chain from past events to a unique event-token.

Jeffares (2008) states that Cleland’s (2002) discussion on the patterns of reasoning between experimental sciences and historical sciences is not complete. Cleland states that historical sciences are not interested in regularities across event-types, but Jeffares claims that historical scientists rely on background theories, which he explains are theories about regularities, to support the causal chain between event-tokens. Jeffares discusses the work of archaeologists Schick and Toth (1993) which isolates the causal mechanisms of marks left of bones. They use experimentation to differentiate between canine marks and tool marks on bone. Jeffares states that when historical scientists employ experimentation they assume that past events produce regularities which determine later phenomena. Historical scientists are conducting repeatable experiments to verify claims, and therefore using patterns of evidential reasoning employed by physical scientists. Therefore, historical scientists have the confirmatory apparatus of the experimental sciences available to them. Jeffares states that historical sciences do not just employ regularities to verify causal relationships of event-tokens, they explore regularities for their own sake. This is because regularities can be thought of in terms of process-types and process-tokens, which are useful for telling us something about the

world. For example, returning to archaeology, a specific incident of bone damage from a canine would have enough in common with any other incident to token of the process. In this case a contemporary experiment reveals a process that is helpful for studying past events. Once a process is understood it can become part of a phenomenon's explanation. In addition, each time a process is successfully used to explain an event-token it gains legitimacy. Jeffares states that the combination of Cleland's procedure of eliminating alternative hypotheses and the confirmatory work of processes provides support for explanations of historical phenomena.

Together, Cleland (2002) and Jeffares (2008) give a comprehensive discussion about the robust nature of research in historical sciences. In spite of this, it is important to realize that historical sciences have been considered by a number of physicists and chemists to be not as robust as the experimental sciences. Cleland shares that Henry Gee (2000), an editor of *Nature* stated that historical hypotheses are not scientific because they cannot be tested. Bias toward experimental sciences has found its way into education. Science textbooks that tend to portray the process of science in a simplified manner by promoting one scientific method that privileges experimental science (Blachowicz, 2009). This claim is supported by Decker, Summers and Barrow (2007) who state the treatment of scientific research in biology textbooks could easily lead students to think that experimentation is the only way to test hypotheses and that experimental science is the only way to investigate the natural world. They acknowledge that experimentation is an important part of biology, but that many biological disciplines that do not lend themselves to experimental studies. Decker et al. (2007) express concern that privileged treatment of experimental science in textbooks will not give teachers

opportunities to present alternative ways of generating explanations about natural phenomena. This bias about experimental science is found in research about NOS aspects. Abd-El-Khalick and Lederman (2000) found that 75% of the participants, who were mainly science majors and pre-service teachers, did not realize that scientific knowledge could be developed without the use of experiments. They refer to Darwin's theory of evolution as a powerful theory that was developed solely on observational evidence. In summary, scientists in the historical sciences often employ different evidential reasoning and methods to explain phenomena and these methods have not always been understood or well received in certain portions of the scientific community. The privileged place of experimental sciences has filtered down into the educational system, and scientific methods such as observations are not always introduced or explained well to students. This in turn effects the type of explanations teachers use to explain phenomena.

Historical sciences often employ a different type evidential reasoning because they ask different questions from experimental sciences. The biological sciences can demonstrate these different types of questions. There are two types of questions that can be asked about biological traits. First, there are 'how' questions which are concerned with how a trait functions, such as;" How does the heart pump blood?" Second, are 'why' questions that ask why a trait exists, or came to exist; such as Why does the heart pump blood?" These question types are directly related to two different types of causes. The first type of cause is a proximate cause which involves the physiology of individuals. The second type of cause is an ultimate cause, which involves the evolutionary history of a species (Mayr, 1961). Evolutionary biology is concerned with questions that involve the

past and this necessitates a specific type of explanation. These explanations must satisfy two requirements. First, they must be concerned with unique events in time and space; and second, these events must have a causal connection between them. It is not necessary to know all the causal events that lead to a later event, but it is important that these earlier events demonstrate a causal connection with the later event (Kampourakis & Zozga, 2008). In summary, as a historical science, evolutionary biology has explanations that rely of a sequence of unique events that involve change from one state to the next. These explanations are challenging for students because evolutionary explanations have properties that are difficult for people to accept and this is because of conceptual barriers (Evans, 2008).

Cognitive Biases

Evans (2008) explains that evolutionary concepts are counterintuitive to people's everyday understanding of the world. People use intuitive reasoning to help them understand everyday experiences. This way of navigating the world is generally successful, but there are three cognitive biases associated with intuitive reasoning that restrict cognition and make evolutionary concepts difficult to learn.

The three cognitive biases that interfere with understanding evolution are a) essentialism, b) teleology, and c) intentionality. Essentialism is the idea that each type of living thing has an essence that makes it unique. This idea is thought to have several useful functions. First, essentialist thinking may help humans view the world as stable and unchanging, which could be useful when learners are trying to understand what is happening in the world. Second, this thinking may help humans classify things, and help learners make inferences about those things. Being able to classify and make inferences

about things streamlines the learning process when learning new concepts. However, essentialist reasoning is in direct conflict with the idea of evolution which states living things change (Evans, 2008).

Essentialist thinking would make it difficult for a student to accept explanations that have change at their core. Evolutionary explanations talk about how species change from some past state to arrive at the presence situation. The essentialist who believes species are stable and unique would have a challenging time believing evolutionary explanations. The second cognitive bias is teleological thinking.

Evans (2008) describes teleological thinking as the human tendency to believe that certain animal behavior is driven by purposeful or goal-directed thinking. Living things that cannot think are simply responding to environmental conditions and internal signals. Teleological thinking may help humans distinguish between living and non-living things. We can detect activity generated from a physical cause, such as a rock rolling downhill, which suggests a non-living thing; vs. one apparently driven by a function, such as a rabbit running down the hill. This thinking may be useful for survival, and it appears early in childhood. Evans states that this thinking is also a barrier to learning evolution.

The use of the term teleological in connection to student misconceptions refers to the intentional actions of an individual organism or external force to move that organism toward a desired pre-determined state of being (Kampourakis & Zogza, 2008). The inclination for students to assign goals to the actions of organisms is in direct opposition to evolutionary explanations that while they display causal relationships, they do not involve the needs or wants of the organism.

It is important to mention that there is another use of the term teleological, and this is in connection to scientific explanations in evolutionary biology. Lennox (1993) states that explanations that involve natural selection are inherently teleological. This means that the presence of a trait in an organism is due to the advantage that trait imparts on that organism, and therefore is selectively-favored. Because the term teleological can be used in two very different ways, both of which may be referring to evolutionary biology, it is important to define which use of the term is being privileged. The third cognitive bias is intentional thinking.

Intentional thinking, according the Evans (2008), is the tendency to assign intentions to actions even when there are none. This thinking assists in achieving human goals and intentions. The barrier to learning evolution is the resistance to the idea that natural phenomena are not intentionally-made or designed. As with teleological thinking, intentional thinking is contradictory to evolutionary explanations which do not include organisms intentionally causing change for their own needs.

Together these three cognitive biases make the counterintuitive idea of evolution very difficult to understand, and therefore evolutionary explanations hard to accept. Students are faced with explaining changing species through non-teleological, non-intentional processes, which contradict the idea of stable organisms whose actions are purposeful and goal-driven. Another aspect of intuitive reasoning is students do not perceive their thinking to be wrong, and this may contribute to resistance of evolutionary ideas (Sinatra, Brem & Evans, 2008). In summary, intuitive reasoning promote misunderstanding about, and barriers to learning evolution by making it difficult for

students to understand and accept evolutionary explanations. One approach to mitigating these cognitive issues is to use conceptual change strategies in the teaching of evolution.

Conceptual Change

Misunderstandings are ideas students hold that do not conform to accepted scientific explanations. Misunderstandings may arise when students incorrectly incorporate new concepts into their existing conceptual frameworks in formal learning settings, and these ideas are termed misconceptions. (c.f. Driver & Easley, 1978). Misconceptions are tenaciously held by students and difficult to overcome (Alters & Nelson, 2002; Bishop & Anderson, 1990; Nehm & Reilly, 2007). However, instructional strategies that explicitly focused on bringing about conceptual change demonstrated some learning gains thought not at desired levels (Bishop & Anderson, 1990; Nehm & Reilly) In addition, misconceptions remained part of the student's conceptual ecology, and seem to be used in specific contexts (Evans, 2008; Kampourakis & Zogza, 2008, 2009) The constructivist perspective on learning states that individuals purposefully and actively construct their own knowledge in an effort to make meaning of the world around them. Knowledge construction happens through personal experiences with the physical world and social interactions. In addition, knowledge construction involves conceptual change, which means the act of learning involves the extension or restructuring of an individual's currently held conceptions. Because learning can be seen as the process in which conceptions change, the students' prior knowledge must be considered in order to promote that change. For this change to take place students must have a chance to express their ideas explicitly so they can reflect upon them. Another important aspect of learning is that it is context dependent, different content domains have their own characteristics

and therefore instructional practices should focus on the specific domain of interest (Driver & Oldham, 1986). There are two important implications of Driver and Oldham's constructivist view for teaching evolution. First, the student's prior knowledge of evolutionary concepts must be considered, and second, instructional approaches should specifically focus on the characteristics of evolutionary explanations, one of which is their historical structure.

Another approach to prior knowledge and its impact on learning comes from Wandersee (1992) who discusses the historical nature of knowledge. He introduces the idea of the historicality of cognition. *Historicity* means being "based on or reconstructed from past events" (Wandersee, 1992, p. 424). Wandersee suggests that a person's ability to access past experiences is necessary for cognition. He states that two important functions of memory are to store information, and to retrieve that information when needed. The meaning a person made of this stored information determines whether an individual's cognitive history contains views that can be considered misconceptions. Wandersee (1992) suggests the use of historical stories in teaching because the structure of stories display historicality, which help learners connect new concepts to existing ones.

If learners could more effectively connect scientific concepts to current concepts, through the use of stories, this could be a promising instructional strategy for teaching evolution. Biology educators Reiss, Millar and Osborne (1999) call for "greater use of one of the world's most powerful and pervasive ways of communicating ideas, the narrative" by recognizing that its central aim is to is to present a series of explanatory stories. Specifically they propose that one of the stories should cover the evolution of species through the process of natural selection. Norris, Guilbert, Smith, Hakimelahi, and

Phillips (2005) talk about the idea of narrative explanations, specifically intrinsic explanations which explain natural phenomena and are part of the scientific body of knowledge. They state that phenomena connected to historical sciences such a paleontology, geology and biology are often unique events that do not reoccur, and therefore the use of narrative explanations would be beneficial when clarifying these phenomena, because retrodiction is necessary in these cases (Norris et al., 2005). Therefore, a narrative explanation can be thought of as; a narrative that unifies and gives meaning to a unique set of causal events. As mentioned above, Cleland (2002) states that explanations in historical sciences have a narrative quality, and therefore the Norris et. al proposal of using narrative explanations for biological phenomena may be a suitable one.

The idea that the narrative form is a powerful way to communicate was not lost on Darwin who used historical narratives to convey his theory of evolution through natural selection. Kitcher (1985), a philosopher of science, claims that Darwin's success in transmitting his theory was due to his use of historical narratives. He calls these narratives "Darwinian histories" (Kitcher, 1985, p. 132). Gould also talks about Darwin's use of historical narratives and how he was an "articulate champion" (Gould, 2002, p.1333) for their use. Gould points out that Darwin argues for the use of unique histories concerning specific lineages to answer questions about current species. Gould continues by stating narratives or histories explicitly tie current states of being to unique past events, and to the invariant laws of nature. Therefore, he claims that the explanatory power of narratives are as robust as any physical explanation, if existing evidence provide a causal link from past events to the current states of being.

The type of narrative described above by Kitcher (1985) and Gould (2002) are distinct from other types of narratives or stories such as those discussed earlier and are used in educational setting. As mentioned, historical sciences must look to the past to answer questions, and the explanations generated from these scientific investigations are historical in nature. These explanations are called historical narratives. At this point, I will only distinguish between historical narratives used by scientists and other narratives which includes stories. In Chapter Two I will be discussing a definition of a story that contains specific structural elements (c.f. Klassen, 2009), but at this point, narratives and stories are interchangeable.

Earlier, I discussed how the structure of evolutionary explanations conflicts with the everyday thinking patterns of learners. These patterns often contain cognitive patterns that promote misconceptions. However evolutionary explanations are similar to the structure of stories, and science educators have advocated for and used stories to transmit evolutionary ideas. The structure of stories may also help learners connect new concepts to old (Wandersee, 1992). So while evolutionary explanations may be difficult for learners to grasp, using stories to convey these explanations appears to be a promising approach. Chapter Two discusses the use of the history of science (HOS) stories in science education, along with the construction and evaluation of science stories.

CHAPTER II

LITERATURE REVIEW

The previous chapter describes the problem with evolutionary explanations and suggests the use stories as a way to deliver evolution content. This chapter builds a case for the use of stories by considering two areas of literature, the use of history of science (HOS) in science education, and the use of stories. This chapter begins by discussing the background and research of HOS in science education. It continues by discussing the advantages of story structure from the cognitive science literature, and the use of stories in science education. The chapter finishes with a discussion of structure and construction of a science story and finally to Klassen's (2009) innovative work on the structural components of science stories, which includes the construction and analysis of a science story.

The first chapter discussed how conceptual change theories can give educators an approach for alleviating misconceptions; which as we saw in Chapter One, is a major barrier to learning evolution. This section of the chapter considers the use of HOS in science education, and how it may mitigate some of the learning issues discussed in Chapter One.

History of science refers to the history of the scientific enterprise. The AAAS (1993) states that science is a human enterprise which seeks to find out how the natural world works, and that HOS is the history of the ideas and activities of this enterprise. Hvolbek (1993) states that HOS includes scientific thought and the discoveries and advances that have come from that thought. Sometimes history is bundled with philosophy of science and it is referred to the history and philosophy of science (HPS);

which is the case when Matthews discusses the role of HPS in science education (1994). While the history of science and the philosophy of science are separate but intertwined concepts, for the purposes of this study which focuses on the history of and not the philosophy of science, the two terms are considered interchangeable. When reporting on the content of a research paper, I will use the author's original terminology. However, when discussing the paper, I will use the term HOS, as defined above.

Another issue with terminology is the myriad of terms used for narratives or stories used as educational tools. As discussed in the first chapter, the use of narratives has two distinct purposes, in the historical sciences explanations of scientific phenomenon are often historical narratives, and narratives in other contexts, such as stories used in educational settings. The research discussed in this chapter uses a variety of terms to indicate a variety of narratives and teaching approaches using narratives. At this point, I will use the authors' own terms, and give their definition if available. Since the research discussed in mainly from science education, I am making the assumption that the narratives/stories mentioned (unless told otherwise) are not historical narratives used by scientists for the purpose of explaining a phenomenon.

HOS in Science Education

Because of the perceived benefits of HOS, there is a long tradition of science historians, philosophers and educators calling for its inclusion in science education (Conant, 1947, 1951, 1957; Duschl, 1985; Matthews, 1994; Monk & Osborne, 1997; Rudge & Howe, 2009 and many others). One benefit in particular that pertains to teaching evolutionary biology is that HOS can help teachers discover and understand conceptual difficulties in their students (Wandersee, 1990a). Even with such staunch

academic defense there has been resistance by teachers to embrace HOS in the classroom Monk and Osborne (1997). One reason given is the teachers' belief that science is an established body of knowledge. History becomes less critical if teachers do not understand the nature of science. Other issues reported by Monk and Osborne are the need to teach science content and classroom management considerations. In such an environment, if HOS is thought of at all, it is as a supplement to add cultural or human interest. If HOS is going to be used more frequently teachers will have to feel it is worthwhile to include. This section highlights a growing body of empirical research that gives support to the efficacy of using HOS for the teaching of science educational goals.

In the mid-1990's two research studies by Jensen and Finley (1995, 1996) demonstrated that the use of historically-rich materials to teach evolution was moderately effective in changing student conceptions. The authors stated that the explicit use of historical materials in combination with the use of conceptual change strategies was a departure from previous studies on the teaching of evolution. The historical materials used in the first study included the evolutionary theories of Lamarck and Darwin, the experiments of Dr. August Weismann and the tradition of human circumcision. The students received outlines for the theories and the teaching assistant (TA) covered these points through a lecture. The TA taught Lamarckian views of evolution before Darwinian evolution. This sequence was to help students express their views on evolution and then challenge any alternative conceptions with historically valid arguments. Weismann's experiments and human circumcision provided evidence against Lamarck's principles. Once all the material was covered, the students worked in pairs to complete a set of tasks involving both Lamarckian and Darwinian theories of evolution. The pairs were required

to come to consensus. The purpose of this last step was to urge the students to explore the disadvantages of Lamarck's theory and the advantages of Darwin's theory. The entire class reconvened for a discussion about the answers. This intervention was covered in a two hour period during one lab. In the first hour of the next lab session the evolutionary theories were reviewed by the TA and once again the students were able to ask questions if needed. The pre and post-test were given one week prior and week after the intervention. The participants in this first study were in a general college biology course and were considered underprepared with diverse ethnic and socioeconomic backgrounds. There was no control group in the study because the researchers wished to maximize the types of conceptual changes observed. The assessment consisted of multiple choice questions in which the students explained their choices, and several short answer questions. The study reported both changes in scores and the sophistication level of the student response. The responses were rated "best understanding" for correct answers that had correct reasons supporting that answer, "functional misconception" for correct answers that incorrect reasons, "correct/incomplete" for wrong answers that were supported by correct but incomplete reason, and "worst understanding" for wrong answers supported by wrong reasons. Results demonstrated that the percent change in correct responses for all questions from pre to post-test went from 23% to 45%. For the of sophistication, only 23% of pre-test responses showed the best level of understanding, and this increased to 45% for the post-test, there was a 98% increase in the number of responses from the pre to post-test. There was also a 65% improvement in students' functional misconceptions. Even with these encouraging results, less than 50% of questions were answered correctly on the post-test. Jenson and Finley (1995) concluded

that because the theory of evolution is complex, and not all the necessary ideas were learned in the available time, the instructional approach used in their first study should have be extended. In addition, this extra time should also be used to deal with specific misconceptions.

The second study of Jenson and Finley (1996) was also conducted in a general college biology course with a group of students similar to those in the first study. This second study was expanded to include six days of instruction in order to incorporate more material. They used two instructional approaches, traditional lecture and paired problem solving (PPS) with mini-lectures. Along with these instructional approaches, they used traditional (TC) and historically rich evolutionary curricula (HRC). The TC curriculum starts with Darwinian evolution and moves to modern ideas of evolution. The HRC curriculum starts with evolution theories before Darwin, including Lamarck, Cuvier, and Paley, and finishes with Darwinian evolution, and it includes information on Darwin's life. This means the study had four different groups; traditional lecture with TC and another with HRC, and the PPS approach with TC and another with HRC. The assessment was similar to that in the first study, it included multiple questions with an explanatory portion and short answer questions. The student responses were classified as Darwinian concept, of which there were six, and Non-Darwinian concept, of which there were four. Each student was given a Darwinian concept score and Non-Darwinian concept score for the pre and post-test. The pre-test Darwinian concept scores for the four groups ranged from 43% -55%. The post-test Darwinian concept score range increased to73% - 86%. The largest positive change in percentage of Darwinian concepts was in the HCR/PPS group with a change of 45%. The smallest positive change was in the

TC/lecture group with a 22% change. So all groups benefited from the various interventions, but the HCR/PPS group had the greatest increase of Darwinian concepts in the responses. This group also had the largest decrease in their Non-Darwinian score. The Non-Darwinian concepts were classified into (a) Teleology (TE) goal or purpose driven change, (b) Lamarck (LM) use or disuse of organs and/or changes passed to offspring, (c) Natural Theology (NT) change attributed to a supreme being, and (d) Other alternative concepts (OAC) conceptions not included in first three categories. All four groups had the largest decreases in TE and LM concepts. The HCR/PPS group having the greatest total loss of Non-Darwinian concepts (45%). Overall, the largest difference in mean scores was between the HCR/PPS group as the most improved, and the TC/lecture as the least. Therefore the researchers conclude that the historically rich curriculum coupled with the paired problem- solving strategy was effective for conceptual change (Jenson & Finley, 1996).

The results of the second Jenson and Finley (1996) study demonstrate that students in the historically rich treatment groups provided more scientific explanations while having a greater loss of non-Darwinian explanations compared to the groups using traditional curricula. The historically rich curricular material used the theories of Cuvier, Paley, Lamarck and Darwin, while the traditional material only used Darwin’s work. This supports the explicit use misconceptions in the form of past theorists/scientists ideas as a way discussing them in a non-judgmental way. One limitation of this study was that the historically rich curricular material was delivered as a traditional lecture or as several mini-lectures. There was no mention that either lecture type used a story structure. So while the results support the use historically rich curricula for the teaching of

evolutionary biology, it does not provide support for the use of stories as a context for the HOS materials. The next paper by Abd-El-Khalick and Lederman (2000) discusses the impact of HOS courses on the learning of NOS.

In a widely cited paper in the fields of NOS and HOS research, Abd-El-Khalick and Lederman (2000) discuss how the use of the HOS to teach the NOS concepts has long been promoted as a solution to the resistance of NOS objectives. They state that in spite of 70 years of science educators arguing that HOS can play an important part in assisting students gain a better understanding of the scientific enterprise, there are no empirical studies that support the use of HOS courses as a way to enhance the retention of NOS aspects. The purpose of their study was to determine if HOS courses have an influence on undergraduate students and pre-service teachers' conceptions of NOS, however Abd-El-Khalick and Lederman caution that the use of historical materials, such historical narratives used in a HOS course may be difficult for students to engage with because their current conceptual framework, in at odds with historical perspectives. They suggest that students should be equipped with a current conceptual framework of NOS prior to enrollment in the HOS classes, which would serve to enhance their NOS understanding with examples and stories. Abd-El-Khalick and Lederman feel that teachers should have a variety of examples or "stories" from HOS as a way to put NOS aspects into context.

Abd-El-Khalick and Lederman (2000) state that the definition of NOS has many variations, and that philosophers, historians, sociologists of science and science educators all may disagree on a specific meaning of this concept. The authors claim that the lack of agreement of a definition is not an issue at the K-12 level because there are generalities

concerning some important aspects of the NOS that are agreed upon. They feel it is these concepts that should be focused on in science classes. The specific NOS aspects that they focused on in this study state that scientific knowledge is (a) tentative, (b) empirically-based, (c) theory-laden, (d) based on inference, imagination and creativity, and (e) socially and culturally embedded. In addition, they feel that the difference between observation and inference, along with understanding theories and laws are important aspects.

The study's research questions looked at (a) the influence of HOS courses on college students' NOS conceptions, (b) if students who enter HOS courses with current understanding of NOS concepts gain an enhanced perception of NOS, which means a prior NOS aspect was enriched (stated more clearly with use of a historical example to support the view), and (c) what HOS aspects, such as course objectives, teaching methods, teacher commitment to NOS and classroom dynamics, impact students learning of NOS concepts.

Abd-El-Khalick and Lederman (2000) classified their study as interpretive with data collection continuing throughout the semester in which the research was conducted. The participants were 181 undergraduate and graduate students, most of who majored in a biological science or general science, and 9% of the students were pre-service teachers. For the study, the researchers sorted the participants into two groups. The first group was 166 students enrolled in a HOS course in a fall term at a west coast mid-sized state university. The second group was the 15 pre-service teachers enrolled in a master of arts in teaching (MAT) program at the same university. Three history of science classes were part of the study; a survey of the HOS course, studies in scientific controversy course and

an evolution course. Only the professor of the evolution course made an explicit commitment to help students learn NOS objectives.

Data in the Abd-El-Khalick and Lederman (2000) study came from several sources. To gauge the students' understanding of NOS, pre and post-instruction open-ended questionnaires were administered in the first and last week of the fall semester. Semi-structured interviews were conducted post-instruction to have the students justify their questionnaire responses. The participants had access to their pre and post-instruction questionnaires. These interviews included a set of follow-up questions that emerged from conducting the pre-instruction interviews. These sources were used to create a profile of the students' NOS views. Data concerning the course instructor and the course itself came from semi-structured interviews with the instructors, and classroom observations, along with the course syllabi. These data sources were used to generate profiles of the course. The profile included instructor priorities, view of HOS and science and the teaching of science, teaching approach and course objectives. In addition to interviews the courses were attended by the primary researcher and audio-taped, field notes were also generated. The objective of the class observations were to document when NOS aspects were emphasized.

The reported results of the Abd-El-Khalick and Lederman (2000) research revealed that the changes in NOS views were slight, overall only 3% of participants had changes in two or more of the seven NOS aspects. Even though overall results amongst all participants in all HOS courses was small, two interesting things emerged. First, students who entered HOS courses with adequate NOS views, had relatively more participants demonstrate a positive change in NOS views. Second, most of the NOS view

changes were in instances where the NOS concept was explicitly taught. The evolution course taught NOS views explicitly and the percentage of change in one aspect amongst the students was 27% for the student group and 40% for the pre-service teacher group. In the Survey course change in one aspect amongst the students was 17%, and in the controversy course change in one aspect was 17%.

Abd-El-Khalick and Lederman (2000) state that the most significant important finding in the study was that HOS courses only had a small influence on NOS views. They state that this study does not give empirical support for the claim that HOS courses improve NOS views. The paper gives four factors that may impede using HOS. First, there appears to be an inherent barrier in HOS as a tool for learning NOS aspects. The authors believe this is not unexpected in light of conceptual change theory. The current conceptual framework of students maybe too incongruent with the framework needed to find meaning in historic materials. The second barrier was the choice of professors to use implicit approaches to teaching the NOS aspects vs. explicit ones. As noted before, virtually all changes involved NOS aspects that were taught explicitly. The third barrier was the students' NOS misconceptions coming into the HOS courses. The fourth barrier may be that the HOS course objectives may not give priority to NOS aspects. With these barriers in mind, Abd-El-Khalick and Lederman suggest that students should be given an adequate framework of NOS concepts before entering HOS courses. This may help students leave the HOS course with enhanced NOS views.

Abd-El-Khalick and Lederman (2000) discuss four implications that emerged from the study. First, HOS courses should use explicit approaches to teaching NOS aspects. Second, explicit approaches may not be enough in the face of misconceptions.

Exposure of the misconceptions coupled with the use of explicitly delivered HOS examples may be more useful than explicit approaches alone. Third, HOS alone will not help learners gain adequate NOS views. Students should have adequate NOS views before entering HOS courses. Fourth, HOS is an established and separate discipline, and it should not be assumed HOS goals align with pre-service teachers. There should be discourse between science educators and historians to help them be more aware of science teachers' needs.

The Abd-El-Khalick and Lederman (2000) paper has a prominent place in science education research, but it is not without limitations. There are two limitations that are pertinent to my study, one in the research design and one in the paper. The first limitation is the use of HOS courses as a strategy to improve student understanding of NOS aspects. In the literature section of the paper, Abd-El-Khalick and Lederman talk about the many years of support by science educators for the use of HOS as a way to improve student understanding of the scientific enterprise. While some of the references such as Matthews (1994) do indeed advocate for coursework in history and philosophy, others such as Monk and Osborne (1997) and Wandersee (1992) talk about the inclusion of HOS in the science classroom. The paper conflates the idea that the inclusion of HOS in science education only means students taking additional HOS courses. The problem with this, is that additional HOS coursework may not be the best way to improve NOS understanding. First, additional coursework may be difficult from programs of study that are already very busy. Second, and more importantly, HOS courses may not be designed for pre-service teachers who were the focus of this study. The HOS courses in the Abd-El-Khalick and Lederman (2000) study were not necessarily aligned with the study's goals,

one of which was explicit teaching of NOS aspects. Only the evolutionary course was taught explicitly. Other issue could be an alignment problem with the very nature of NOS itself. This study was focused on a list of seven specific NOS aspects and the HOS courses may have had a more global framework for the NOS, or in the case of the evolutionary course, talked about NOS aspects outside the list, such as an appeal to a supernatural power. Since NOS aspects not on the list of seven would not have been looked for, there would be no awareness how other aspects of NOS were being covered in the courses. Because it is very likely that the HOS courses emphasized other NOS aspects besides the stated seven, the low scores reported may be due in part of the narrow definition of NOS.

The second limitation is concerned with the paper's lack of clarity on the type of historical materials used in the HOS courses attended by the students in their study. At different points in the paper the HOS course curricular material is referred to as historical material, and historical narratives. In additional the paper makes the recommendation that teachers should have at their disposal HOS "stories" as a way to place NOS aspects in context. With no specific definition of what a historical narrative is versus a "story", the reader is left wondering why the use of HOS in the science classroom is desired, when its use may not be beneficial to college students including pre-service teachers. This clarity is especially important in light of Abd-El-Khalick and Lederman's (2000) caution about the use of HOS. The paper uses the example of Kuhn's *The Essential Tension* (1977), in which Kuhn describes his difficulty considering Aristotle's writings on motion because of his own Newtonian framework. Until Kuhn was able to find an alternative way to think about Aristotle's work he was not able to understand it. In Abd-El-Khalick and

Lederman's (2000) example of Kuhn's (1977) difficulty with Arisototle's writing, they imply that the historical materials used in the HOS courses could be of this caliber. However, the reader is not privy to the type of HOS materials so there is no way of determining if there was a lack of gain in understanding NOS aspects because of the material, or for some other reason.

Abd-El_Khalick and Lederman (2000) on the one hand appears to advocate for the use of HOS as a way to improve student understanding of NOS aspects, but also feel there are inherent issues to using HOS in the classroom. There appeared to be no attempt to alleviate this issues because the research was limited to having students taking additional coursework in HOS, which did not seem aligned to the needs of the students, or the researchers' educational goals. While the research results were not significant, the evolutionary HOS course, which was taught explicitly in reference to the NOS aspects, had results that were more promising than the other courses. So while this paper exhibits a tension about the use of HOS and had limitations in the design, the results were encouraging enough to prompt further research, both on the use of HOS and the explicit delivery of NOS aspects in the classroom.

The results and limitations of the Abd- El-Khalick and Lederman (2000) paper gave a direction for further science education research. As mentioned earlier, they felt that the use of historical narratives may have impeded the effectiveness of HOS in science education. Rudge and Howe (2009) continue the discussion on the use of HOS to teach NOS aspects. They discuss the limitations of using complete historical narratives, in this case, historical vignettes as used by Monk and Osborne (1997). Instead they

propose an alternative, instrumental approach to the use of the HOS. They describe both a HOS lesson plan and the empirical research study conducted to determine its efficacy.

Rudge and Howe (2009) have two main purposes in their paper titled: An explicit and reflective approach to the use of history to promote understanding of the nature of science. First, they review and critique the Monk and Osbourne's (1997) model for using the history and philosophy of science in the science classroom. Second, they discuss their lesson plan and research of their own use of HOS in a biology course and how it improved the students' understanding of certain NOS aspects. Rudge and Howe state the purpose of the critique is to discuss the weaknesses of the Monk and Osbourne approach and to introduce an instructional method that is more aligned with Monk and Osbourne's own rationale for including HOS in science instruction. This rationale is based on the idea that students bring their own conceptions to science classes and the inclusion of explicitly constructed knowledge about the products and processes of science through the use of HOS will the students identify and refine previously held ideas about the nature of the scientific enterprise.

Rudge and Howe (2009) report that Monk and Osbourne (1997) list three advantages of using HOS. First, by using highly contextualized historic examples the students will implicitly learn about nature of science aspects. Second, students engage in critical thinking about scientific evidence and will gain new knowledge about nature of science. Third, the use of historical vignettes will help teachers lacking knowledge in history of science, to be perceived as authorities in this area. They discuss the strengths and limitations of the Monk and Osbourne approach by pointing out that their method does not support the three advantages listed by them. First, the Monk and Osbourne do

not introduce the historical vignettes to discover the students' misconceptions, but to add another viewpoint. Rudge and Howe think this is unusual considering Monk and Osbourne's commitment to constructivist learning theory. Second, the reading of historical vignettes of others' experiences could become a passive activity to the student. Third, the detail of the historical materials is too deep and may be considered excessive by the students. Fourth, the introduction of the modern scientific view may convey the idea that previous investigators were not acting in a scientific way, and that the studying of historic efforts may be conceived as fruitless.

Rudge and Howe (2009) give a detailed account of an instrumental approach to using the history of science to teach content and nature of science objectives. This approach invites the students to experience a phenomenon in a similar sequence as the scientists they are studying. In this way students can overcome the misconceptions in the same way the scientists did. The example used in the paper is an eight-day unit on scientific research that led to the understanding of sickle-cell anemia. On the first day of instruction the students are introduced to the disease as a mystery by presenting them with the symptoms seen by Dr. Herrick, who is one of the historic figures in this lesson plan. Each day the instructor gives the class a new set of evidence and asks guiding questions to help the students move toward solving this scientific mystery. Students are asked to journal outside the classroom on their in-class experiences, areas of confusion and ideas for further inquiry. Along with content questions, nature of science probes are also given to the students to consider in their journals. Rudge and Howe (2009) stress that is process is designed to be explicit and reflective and this supports previous research that found that this technique improves student learning of NOS aspects

Rudge and Howe (2009) conducted their research at Western Michigan University in a BIOS 2700 course, which is Life Science for Elementary Educators II. The students in this class are pre-intern students who are planning to be pre-secondary teachers. The pedagogical approach in this course is inquiry-based. Most of the students do not have a strong foundation in math and science.

The students were given a modified VNOS survey, which is a survey on the students' view of nature of science aspects. Along with the survey, students were randomly chosen to participate in an interview conducted by the second author. The purpose of the interview was to validate the survey instrument and investigate any irregular answers. At the end of the unit the students took the survey again; and another group of randomly chosen students were interviewed. This data was used to develop profiles of pre and post-unit nature of science views. This was accomplished by placing the segments of meaning into categories. The data was reduced by multiple code applications, and this resulted in individual and aggregate profiles. These profiles were analyzed by the second author to find instances where NOS views changed considerably between pre and post-instruction. It was specifically noted when students revealed that the unit on sickle-cell anemia influenced their change in NOS views.

Rudge and Howe (2009) stated that the study demonstrated a considerable number of students did have significant change in their NOS views. In addition, some students held more informed NOS views and supported them with examples from the unit. They felt that another encouraging result was that the students demonstrated transfer by answering VNOS questions that were not specifically concerned with sickle-cell anemia. The results were compiled in a table that showed students demonstrated

enhanced understanding of four NOS aspects. The percentage of students that demonstrated change ranged from 12 and 38%. The percentage of students that demonstrated change of NOS views, and provided answers with support ranged from 9 to 15%.

Rudge and Howe (2009) conclude that their approach in using the HOS in an instrumental way to teach both content and NOS objectives is supported by two things. First, their lesson plan as described in this paper, does not have the pitfalls inherent in the Monk and Osbourne (1997) approach. Second, Rudge and Howe argue that the results of the VNOS and interviews demonstrated the efficacy of this approach with sickle-cell anemia. They also stated that future studies are needed to ascertain the effectiveness of other historical examples with this instrumental approach.

The research portion of the paper had a couple of limitations. First, Rudge and Howe (2009) did not explicitly state their theoretical framework when discussing the research. However, they did discuss HOS in reference to being a way to handle student's alternative conceptions. This does imply that their theoretical stance is constructivist. It would have been stronger to state this when introducing their own teaching approach, or discussing the research. Second, while the paper did give ample detail about the research design and that the survey instrument was checked for validity, there was no discussion about inter-coder reliability or controlling for other forms of bias. Third, they did not consider limitations of their own design. One possible limitation is that the research involved only one section of the BIO 2700 course. It would have been good to see if the same improvement trend occurred in another semester, and to have the treatment class have a control class to compare it with. Finally, Rudge and Howe did not provide quotes

from the VNOS or interviews to support their findings. The inclusion of student quotes would have made a stronger argument.

The Rudge and Howe (2009) paper's main purpose was to introduce a HOS teaching approach that considered the best of Monk and Osbourne's (1997) approach and attempt to overcome what they believed to be the limitations of that earlier approach. The authors presented a careful summary and critique of Monk and Osbourne's work. They also provided a robust argument for their approach and a detailed description of their lesson plan and its application. Finally, Rudge and Howe conducted a useful research study on the efficacy of their approach and reported its encouraging results. Even with the limitations of the study, it positively adds to the body of research on conceptual change using an explicit and reflective HOS approach.

The Rudge and Howe (2009) research study adds support to the use of HOS in the classroom. The course materials are modified versions of primary materials that are applied in the classroom as a mystery to be solved. The students move through the lesson and are exposed to more of the story until there is resolution of the mystery. The Abd-El-Khalick and Lederman paper (2000) and the Rudge and Howe paper both talk about using a historical story to help teach NOS aspects. Another use of an historical story is discussed in Sobles and Traver (2003), which talks about the use of HOS to help alleviate negative student attitudes toward science.

The Sobles and Traver paper (2003) is concerned with the increasingly negative public image of science beginning during the last years of the twentieth century. The authors first review the different origins of these negative opinions and then they discuss

their research study on the use of HOS to help modify these opinions. The main portion of their study analyzes student knowledge of the scientific enterprise and how this knowledge influences student attitude.

Sobles and Traver (2003) discuss specific philosophical stances that include anti-science components. The first group is the relativistic science philosophers. This group believes that science cannot be a privileged way to discover how the world works because there is no rational way to choose between alternative epistemologies. The second group is the strong program sociologists who believe since science and social organizations influence each other there is a danger of groups such as Nazis who use science in negatives ways to oppose or harm others. A third group are mystical groups such as creationists who often are opposed to science. A fourth group are pacifists and/or environmentalists who oppose science for its tendency to supply technocrats with materials that can be destructive. Along with this group, critical feminists and others oppose the Western male power base associated with science and those who oppose globalization have expressed disapproval of science. Against these groups are those who believe science as the only true way to gain knowledge and the authors refer to this as scientism. The authors argue that these groups produces tension and this coupled with the slanted and scare media coverage given to science, creates a negative public view of the scientific enterprise. Surrounded by this negative atmosphere, students, who are voluntary at the secondary level, also have to endure poor instructional methods and materials. The authors claim that students in this environment will be more apt to drop out of the pipeline. They also claim that this loss of students in the physical sciences may be minimized by including history of science instruction in science courses.

Sobles and Traver (2003) hypothesis is that physics and chemistry taught without proper instruction in HOS taught in a science, technology and society (STS) context will lead to an inadequate view of science and those views will contribute to poor attitudes toward science. The authors concluded two separate pieces of research to test this hypothesis. The research study's subjects were secondary Spanish physics and chemistry students aged 15 -17 years old. The authors developed a lesson plan using the following criteria: The HOS had to be a storyline, with active teaching techniques and activities included in the instructional methodology. The material has to have HOS as the basis for the instructional material. The materials had to (a) demonstrate how the process of acquiring scientific knowledge actually operates with both grand achievements and problems, and that knowledge is an accumulative process involving many people, and (b) show the hypothetical and tentative nature of scientific knowledge, the limitations of theories, and that unanswered questions still exist. In addition, the materials should demonstrate the contributions of women, discuss the sometimes controversial nature of scientific research, the way science is employed to solve difficult problems, and finally how science can combat issues of pseudo-science. The authors discussed how the criteria were to be included into the classroom setting.

Sobles and Traver (2003) stated that the materials would be presented as a storyline so that students can reconstruct the historical episode and learn the pertinent information as they move through the lesson. The historic ideas and thoughts that scientists held and are parallel to current student's alternative conceptions would be explored and considered while presenting the lesson. They discussed that the storyline

methodology was important because they had no more time to present added curricular items, so the content knowledge had to be presented within the historical storyline.

Sobles and Traver (2003) administered three different survey instruments to randomly selected students from the top three grades in high school. The students in traditional classes which only had a superficial treatment of HOS were considered the control group. The students in classes that used the new curricular materials developed by the authors were in the treatment group. The three surveys focused on different outcomes. Form B was designed to see if students understood how science operates within the discipline and if its achievements have social implications. Form C was designed to see if the student understands the evolution of science and the achievements of scientists and the contributions of Spanish scientists. Form D was designed to gage the attitudes of students about science and scientists.

Form D was given to the students first so the questions from the other surveys would not bias the student's answers to Form D. Two sample groups from the traditional classes were given Forms B and C & D. Both the control and treatment group used a class period in the middle of the last term in the school year to fill out all three surveys. The surveys used open-ended questions. The results of the research demonstrated that the students in the treatment course reported a more appropriate view of science than the students in the traditional classes. The difference between treatment and control groups was statistically significant (a) for all five questions in Form B, (b) for most of the questions in Form C except for questions C4 (Spanish scientists) and C7 (obstacles in scientific practices), and (d) for all three questions in Form D. Sobles and Traver (2003) conclude that when HOS is incorporated explicitly in the classroom students learn

appropriate views about the science enterprise. They make a claim that when students have an appropriate scientific view they will harbor an appreciation for science and scientists and therefore will not have negative attitudes toward science. The significant changes demonstrated in the surveys from the control to the treatment group support this claim.

The main point of the Sobles and Traver (2003) paper was that negative public views of science coupled with poor instructional practices in Spanish secondary schools contributed to students views of science and therefore the loss of potential science students. Their HOS storyline application did demonstrate significant student gains in NOS aspects and improved views toward science as an enterprise. This is quite important when teaching evolution, which as previously mentioned, is a subject that has a lot of misconceptions and negative attitudes attached to it. A limitation of the Sobles and Traver's (2003) that concerns my study is the lack of description of the curricular material, including the storyline. They did give general criteria for the materials, but did not talk about the specific scientists and episodes used, and the nature of a storyline. I cannot assume that the use of the word "story" implies that the historical account being told to the students has a specific structure.

The Sobles and Traver's (2003) storyline approach is similar to the Rudge and Howe (2009) approach, in that the students moved worked through their ideas and that of past scientists with the goal of learning about the process of science. They differed in the historical focus of the lesson. Sobles and Travers concentrated on scientists and their accomplishments. On the other hand, Rudge and Howe focused on the scientific phenomenon of sickle-cell anemia to deliver the course objectives. The focus on a

scientific phenomenon when using HOS, is not unusual. Irwin (2000) uses the concept of the atom to teach both NOS and science content goals. Matthews (2001) uses pendulum studies through history to teach NOS aspects. Klassen (2010) uses the history of the photoelectric effect to teach about the concept of the photon.

Marquee species. In the discipline of biology, a particular species can be used as the representative of a particular phenomenon to deliver both NOS and content goals. Clary and Wandersee (2008) talk about the use of unique species as a way to grab students' attention. They explain that if a unique or rare species is being used to introduce the science curriculum, it is considered a "marquee species". These marquee species are meant to elicit interest and curiosity with the intent of drawing students into a more meaningful study of the topic. A marquee species could have a unique physical characteristic, an important place in scientific history, be of local interest or be attached to a famous scientist. The use of a marquee species has the potential to act as a gateway for the student to build on their current knowledge and incorporate new material (Mintzes, Wandersee, and Novak, 1998). One such marquee species that has been used as a way to introduce and teach evolution through natural selection, this is the moth *Biston betularia*. This moth was extensively studied because of its connection to industrial melanism, a phenomenon that demonstrates natural selection. Industrial melanism (IM) refers to a rapid increase in the frequency of the melanic individuals in many moth species. The dark coloration was first noticed during the mid-late nineteenth century in Britain and later in Continental Europe. The dark moths were first found in the vicinity of manufacturing centers and the increase in frequency of these individuals was believed to be a consequence of large scale air pollution. The investigations of the peppered moth

became widely adopted in biology textbooks as an example of natural selection (Fulford & Rudge, 2016; Rudge & Fulford, 2011). The moth, in its role as a marquee species and the phenomenon of industrial melanism create an exemplar that when delivered with the use of HOS can help students learn about the NOS and natural selection.

HOS exemplar. Michael Majerus was a well-known geneticist in the field of evolutionary biology, who wrote the definitive book on melanism; *Melanism: Evolution in Action* (1998). Majerus (1998, 2005, 2009) was also a science educator who thought that the concepts of natural selection, and the nature of science, specifically, the practice of science and the development of scientific knowledge are illustrated well by the phenomena of industrial melanism. Rudge (2000) also advocates for the use of industrial melanism as a way to teach evolution through natural selection. One approach to teaching with the peppered moth is the instrumental use of HOS that Rudge and Howe (2009) used with sickle-cell anemia (discussed earlier in this section). The instrumental approach uses HOS to introduce a mystery about the scientific phenomenon, and students work through their own ideas and those of past scientists to develop the historical account through multiple class sessions. The students recognize that the ideas they generate are similar to those of past scientists and in this way they are able to share without feeling any judgment. This instrumental approach was used in an evolution unit to explore the phenomenon of industrial melanism in the peppered moth. In a study involving pre-service teachers in an introductory biology course, Rudge, Cassidy, Fulford and Howe (2014), report on the research results on this two week lesson as it relates to NOS objectives. This research study also reported positive gains in students' understanding of NOS aspects.

There are two main limitations to the body of research on the use of HOS. First, there has been an emphasis on NOS research over the last decade in science education, and often research involving HOS is focused on NOS (Abd-El-Khalick & Lederman, 2000; McComas, 2008; Rudge and Howe, 2009; Rudge et. al, 2014). Abd-El-Khalick and Lederman (2000), Sobles and Travers (2003), Rudge and Howe (2009), and Rudge et al. (2014), do mention that the lessons involved with the research studies have content goals; however these papers do not discuss whether or not these goals are met. There are a few of research studies that explicitly discuss the impact of HOS on both NOS and content goals. Kim and Irving (2010), report on their project that looks at the efficacy of a HOS-based instruction on genetic content and NOS objectives. They report NOS goals were better with HOS instruction, but content goals were no different with the traditional lesson, vs. the HOS-based lesson. Clough (2011) also discusses the need to consider science content when using HOS to enhance NOS instruction. As stated above Monk and Osborne (1997), caution that teachers worry that HOS-based curricula short-change science content, and Heilbron (2002) warns that the history in the HOS-based lesson should never be privileged over content. Therefore, it would be a good idea to explicitly discuss content goals in any paper that discusses HOS and NOS.

Another issue with the research discussed in this chapter section, is the lack of definitions for the terms used to describe the delivery of HOS. Abd-El-Khalick and Lederman (2000) talk about historical narratives and stories. Sobles and Traver (2003) discuss the use of historical storylines. Rudge and Howe (2009), and Rudge et al. (2014) discuss mysteries and stories, and Monk and Osbourne (1997) talk about historical vignettes. All these ideas appear to be related to stories and there is a general sense that

the reader understands what this means, but these terms need to be quantified so the concepts attached to them can be discussed and considered in a research setting. The lack of definitions, makes it difficult to ascertain the efficacy of any particular narrative because the basic structure and components are not known or agreed upon. Also, these studies vary in the way the narrative is delivered. Researchers caution that using a complete, stand-alone historical narrative that does not allow the student to reflect on the connections between science in the narrative and current science ideas (Abd-El-Khalick, 2000; Rudge and Howe, 2009). So while the use of narratives seems like promising avenue, this chapter section reveals some gaps in the research that must be addressed. The next section of this chapter in this review attempts to do just that, with a discussion on the advantage of stories and their use in science education.

Narratives

There is a close relationship between the use of HOS and narratives, and this may be due to the inherently narrative structure of historical accounts (Kampourakis & Zogza, 2008). In the case of HOS it seems intuitive to deliver science content in a narrative format. In science education there has been an increased call for the use of narratives, in part due to the perceived benefits of narratives (Arons, 1989; Wandersee & Roach, 1998; Metz, Klassen, McMillan, Clough & Olson, 2007; Stinner, MCMillan, Metz, Jilek & Klassen, 2003). This chapter discusses what is meant by narrative, the support from cognitive science for the usefulness of narratives in learning, and the use of narratives in science education. Discussions of narratives span many disciplines and the terminology in each discipline can have its own specific meaning; narratologist Barbara Herrnstein Smith, brings some clarity to the definition of narrative, she states: “we might conceive of

narrative discourse most minimally and most generally as verbal acts consisting of "*someone telling someone else that something happened"* (Herrnstein Smith, 1980, p. 232). At this point, this definition is useful and the terms narrative and story will be used interchangeably because this is the convention in the literature.

As discussed earlier, various science educators discussed the use of narratives as a way to deliver science content. In Chapter One, Kampourakis and Zogza (2008) advocate for the use of historical narratives as a way to deliver evolutionary explanations. In this chapter, Abd-El-Khalick and Lederman (2000) discusses inherent problems associated with the use of historical narratives. Also in Chapter Three, Rudge and Howe (2009) discuss Monk and Osborne's (2007) use of historical vignettes, and their own use of a mystery and story, and Sobles and Traver (2003) talk about their use of historical storylines. In all of these papers, with the exception of Rudge and Howe, the terms were not specifically defined. The Rudge and Howe paper included Monk and Osborne's definition of historical vignette, which is a "short narratives that place the past scientists work in a rich historical context" (Rudge & Howe, 2009, p. 564). Sobles and Traver, and Rudge and Howe did describe their application of the historical material. However, they did not define the actual terms historical storyline, mystery or story. In science education the terms historical case studies (Conant, 1947, 1951, 1957), or case study are also used (Herried, 1998). Herried defines a case as a story that delivers a relevant message (1998). At this point in the chapter, all these terms are interchangeable with narrative and story. Later in this chapter, the term story takes on a specific meaning and this will be discussed at that point.

Narratives and learning. There is a sense that stories are a privileged way for people to learn, in that stories are a cultural universal, a basic way in which we make meaning out of our interactions with the world (Egan, 1986). Edward O. Wilson (2002) in his editorial, The Power of Story, says that humans make sense out of the world through narratives and that in contrast to stories, the scientific enterprise is distinctly difference from everyday experiences, and therefore more difficult to understand. Wilson argues that if science is delivered as a narrative it will make the content enjoyable and relatable without compromising the nature of the phenomenon being presented. As discussed in Chapter One and earlier parts of this chapter. Educators and researchers have called for science content to be delivered by stories (Aarons, 1989; Kampourakis & Zogza, 2008; Reiss, Millar and Osborne, 1999; and many others).

***Cognitive science*.** The field of cognitive science provides support for the use of narratives. Research suggests that people incorporate narrative structure into their cognitive landscapes and this aids memory and recall. (Bartlett, 1932; Mandler & Johnson, 1977; Thorndyke, 1977). Mandler and Goodman (1982) talk about how people integrate new information about a specific topic, and that a person's understanding of this topic is their conceptual ecology. They explain that stories have structures that people incorporate into their conceptual ecology, which are known as story schema. This schema is a record of the regularities learners discover about stories through interaction with them. It is the regularities in stories that give them a specific structure. Research demonstrates that a person uses their knowledge of this canonical structure to process information, and that this structure improves reading and recall.

Another area of research in cognitive science demonstrates that when people are reading narrative texts, they are able to create inferences that help make meaning of the ideas in the text (Black & Bern, 1981; Potts, Keenan, & Golding, 1988; Millis & Graesser, 1994). It appears that the comprehension mechanisms recruited by the mind when information is presented as a narrative is closely related to everyday experiences, and therefore more natural than the mechanisms used to incorporate other modes of discourse such as expository text, or argumentation. There appears to be stable comprehension mechanisms across people in and between cultures, and these mechanisms process a variety of narratives. For example, the mechanism that provides meaning to the narrative's plot is similar for all narratives from folktales, short stories, oral narrative, graphic novels and film. These mechanisms may render narratives easier to remember than other modes of discourse, which might account for its privileged status over other modes (Mandler & Goodman 1982). There is a body of research comparing expository text and narrative text. In a study on high school students with learning disabilities had more trouble with reading fluency and comprehension when reading expository text. However, the type of question asked impacted the results for reading comprehension in expository text (Saenz and Fuchs, 2002). Another study used EEG data to determine how bridging inferences are impacted by narrative vs. expository text. The research reveals that the participants were able to create bridging inference easier with expository text. The results are inconsistent with other research, and the fact that the expository text was rewritten in a logical and complete piece of text about general knowledge, and not difficult science related information, may have influenced the results (Baretta, Tomitch, MacNair, Lim & Waldie, 2009). Wolfe and Mienko (2007)

investigated the impact of narrative and expository text on learning and memory. This study did not find any difference between the two types of text on learning and memory. However, when the participant's prior knowledge was considered differences in comprehension were revealed between expository and narrative text. Wolfe and Mienko said that sequential expository text had a greater influence on higher knowledge participants, and narrative text had a greater influence on lower knowledge participants. They also state that expository and narrative texts are processed differently, and this means that expository texts prompt a greater integration of content with prior knowledge and narrative text prompt more focused attention on the specific events in the text. Wolfe and Mienko conclude that one text type may not be superior to another, but that the process differences in the discourse modes and a student's prior knowledge may be exploited by educators to improve learning.

While research results on expository and narrative text have reported contradictory results, it remains true that the use of narratives have demonstrated positive results. The important question may not be is narrative text better than expository, but what unique advantages do narratives bring to education. In science education, there are other goals besides those mentioned above, such as the need to engage students, and increase curiosity, to make the content relevant, and to explain science concepts. The next section of this chapter focuses on the use of narratives in science education.

Narratives in science education. As mentioned in the introduction of this chapter, science educators have called for the use of narratives. In the early 2000s science education researchers renewed the conversion about the use of narratives. This discussion focused on the creation and application of historically-based narratives. Stinner,

McMillan, Metz, Jilek, and Klassen (2003) advocates for the use of a contextual instructional approach for all age levels. They discuss different units of historical presentations and the application of these units. The smallest unit is a vignette, they refer to Wandersee (1992) who discusses vignettes and claims are useful for motivation and encouragement. The next unit is a case-study, which are an instructional approach which employs a story-line that delivers the main idea through a group of experiences. A story-line is an account that employs people involved in conflict; which is resolved through the application of the case-study. They mention simple science stories, but offer no definition (Stinner et al., 2003). It appears that vignettes, simple science stories and story-lines are forms of narratives, and case-studies are a way to deliver narrative accounts. Given the different definitions of narratives within science education, as mentioned in the terminology section, or the lack of definitions, what is missing from the conversion is a theoretical framework that would give researchers a working definition of narratives and a direction for future research.

Theoretical framework. In the mid-2000s science education researchers began to tackle the task of clarifying the definition and structure of a narrative. The Norris et al. (2005) paper puts forth a theoretical framework for narrative explanation. This framework defines narrative explanation and its structural components called narrative elements. The purpose of the Norris et al. paper is to discuss the rationale for a theoretical framework involving narrative explanations. The authors point to the literature that advocates for the use of narratives in science education, because of the presumed benefits of these narratives. These benefits include, reducing scientific stereotypes and distortions, increasing student motivation and enjoyment, making science more relatable, and for

engaging students in ethical discussions. The paper's literature review also noted that some science educators consider narratives to be a powerful way to transmit ideas in a meaningful and integrated fashion. The authors' focus is on "the explanatory role of narratives" (Norris et al., 2005, p. 536).

Norris et al. (2005) realize while many empirical issues concerning narrative explanation remain unresolved; they argue that a framework is needed to provide researchers with a way to differentiate between narrative explanations and other types of explanations. Without a framework, researchers would not be able to test their claims about the advantages of narrative explanation. A framework would also provide a more fully developed concept of narrative explanation, which will have merit in science education, and it will help researchers determine the existence of the narrative effect which is purported to enhance memory, interest and understanding.

Norris et al. (2005) define the concepts of narrative, explanation and narrative explanation. This is important because without establishing narrative criteria, alleged examples cannot be identified, and the efficacy of narrative explanation in science education could not be tested. Central to the authors' definition of narrative are eight elements that they believe distinguish narratives from other modes of discourse. The eight narrative elements were chosen from narrative theory and Norris et al. (2005) began with Barbara Herrnstein Smith's quote that at its most basic a narrative consists of "*someone telling someone that something happened"* (Herrnstein Smith, 1980, p. 232). The authors were concerned with written accounts so their translation of the Herrnstein Smith quote is: (a) "someone telling" is the narrator, (b) "someone" receiving is the reader, (c) "something happened" are events and past time. They added the following

elements discussed by other theorists: narrative appetite, structure, agency and purpose. Each of the eight narrative elements is discussed in length.

The first narrative element is called an event-token (something happened). There ample agreement that a narrative must have a sequence of events. Within this sequence, the individual events must be related to one another. This relationship must demonstrate how earlier events have essential contributions to future events, but the relationship does not have to be causal. In a narrative, the importance of earlier events is revealed by the later events. It is only the end of the temporal sequence that demonstrates which earlier event started the entire process. The second narrative element is a narrator (someone telling). Narratives must have someone transmitting the account. It is the narrator that gives meaning to the sequence of events by creating an account that is unified. Narrators determine the point the narrative by choosing the sequence of events they deem best to relate the account, this means the role of the narrator involves interpretation. The sequence of events can help the narrator create suspense and anticipation, which are devices that help make the narrative interesting. The third element is narrative appetite (wanting to know what happened). For a reader to benefit from a narrative, they must choose to engage with and remain connected to the narrative. This is accomplished through narrative appetite. Devices that create narrative appetite are suspense, missing information, tension between events and the inability of the reader to predict what will happen next. Not knowing will happen creates anticipation and expectations of what could happen based on earlier events. The fourth narrative element is time (the past). Time is an essential component because narratives are concerned with the past. Time is also a device narrators can use to make the narrative interesting. The event-tokens do not

need to be told in chronological order, it is up to the narrator to construct a sequence that they believe will create an engaging narrative. The fifth narrative element is structure. Narratives involve events that are contextualized by time. Minimally, there is always a beginning, middle and end time period, distinguishable in all narratives. This can also be viewed as a beginning situation that leads to some change which is followed by a new situation. How a narrator chooses to order event-time helps the narrative maintain the reader's interest. The sixth narrative element is agency. This element is not considered essential by all narrative analysts; others however, start that the basic narrative form must have characters living in a certain context while moving through time. This means narrative events must have an agent of action. Agents, which do not have to be human, experience and may cause unified temporal events. The seventh narrative element is purpose. The purpose of a narrative in the broadest sense is to help the reader understand something about the world. From a constructivist view, it is to incorporate unfamiliar information into a familiar frame. Narratives do this by communicating knowledge through actions, emotions and values. The eighth and final narrative element is the reader (someone receiving). The reader must be engaged in the narrative because to find meaning, the reader must interpret what they are reading, and this requires effort. Readers use a variety of processes to interpret the text, such as inference, hypothesizing, and anticipating. Readers also use their prior knowledge about modes of discourse to help them figure out where the text going. This means if a reader recognizes the text as a narrative, they will react to it in a certain way, like anticipating the next event. This is because they have encountered narratives in the past and understand the structure.

Norris et al. (2005) state that the eight narrative elements are not the only components in a narrative, but that their presence or lack of it gives an account a degree of narrativity. They argue that narratives fall on a continuum. In addition to this narrativity, they believe there is a hierarchy of importance amongst the eight narrative elements. They believe event-tokens, past time and agency are of primary importance to a narrative. They call for research on accounts that lack elements such as agency to determine if they still produce a narrative effect. They call accounts that mixed narrative with non- narrative passages hybrids. The next section in the paper discusses the idea of an explanations and narrative explanation.

The authors admit that the concept of explanation is difficult to clarify and that no one account is satisfactory. They state that explanations at their most basic are intended to make something understandable. Past this, explanations can be classified by their functions, which are, (a) to clarify a concept, (b) to justify why something was done, (c) to describe what was happening, and (d) to describe why something was happening. Explanations can also be classified by type, and Norris et al. (2005) state that they are most interested in scientific explanations, of which they describe five. The first is the deductive or deductive-nomological model (D-N). These explanations must have a covering law that accompanies it, meaning by what laws and because of what prior conditions, does this phenomenon occur? The second type is based on statistical probability. Statistical explanation demonstrates that a phenomenon is probable based on facts and statistically general laws. The third type of scientific explanations are function explanations, which answer questions such as “Why do humans have kidneys?” The fourth is the genetic or historic explanation. This explanation recounts a sequence of

events in which an earlier state was transformed into a later state. The sixth is the pragmatic explanation. This type of explanation answers a specific why question, such as “Why do maple trees lose their leaves in the fall?” With the concepts of a narrative and explanation defined, the next task of the paper was to discuss narrative explanations.

To discuss the concept of a narrative explanation Norris et al. (2005) first make the distinction between phenomena that is atemporal, which means that future behavior was independent of the history of the system. For example, phenomena in classic thermodynamics display time but not history, therefore explanations for these phenomena do not need to utilize narrative structure. This type of phenomena, those that are repeatable and generate predictions, have event-types vs. event –tokens. An event-type is a class of events, where an event-token is a member of the class. On the other hand, phenomena connected to historical sciences such a paleontology, geology and biology are often unique events that do not reoccur; therefore these events are event-tokens. The use of narrative structure is beneficial when clarifying these phenomena, because retrodiction is necessary in these cases. Narrative explanations must contain narrative elements, of which event-tokens and past time are primary, and that explanatory portion appears to be causal. Therefore, a narrative explanation can be thought of as a narrative that unifies and gives meaning to a unique set of causal events.

Norris et al. (2005) are also interested in the presence of a narrative effect. The authors state that their theoretical framework is not dependent on the existence of a narrative effect, but it is of interest because such an effect would be beneficial to science education. This effect, is supposed to increase student engagement, memory and comprehension. The authors believe there is moderate, support for the existence of a

narrative effect, but the nature of this support is general. Research has demonstrated narrative passages are read quicker, understood better, and the information is deemed more plausible and persuasive. It has also demonstrated enhanced memory and increased student effort. The authors caution that while there is support for a narrative effect, their literature review contained about 200 papers on narrative, 45 of which are empirical, but only 26 considered narrative and expository text. However, 13 of these supported the existence of a narrative effect, three favored expository text, and one using the same text an earlier research project found in favor of narrative. The authors believe other variables may be more important than the mode of text, these could be a student's prior knowledge, motivation, reading proficiency and strategy, and activities such as discussion, summarization and cued recall.

The Norris et al. (2005) paper develops framework for the structure of narrative text and produces a well-defined list of narrative elements. It describes specific types of scientific explanations such as the historical or genetic explanations which are narrative in nature. Their discussion of historical explanations and the historical sciences are tied together well, and this leads to their definition narrative explanations; which are narratives that provide scientific explanations to a unique set of events. It is important to note that Norris et al. (2005) creates a valuable starting point in our theoretical understanding of narratives with the compilation, justification and definition of the eight narrative elements. However they only discuss narrative explanations. As mentioned earlier, Stinner et al. (2003) talks about other forms of, and purposes for narratives. Metz et al. (2007), continues the conversion on narratives for science education.

Metz et al. (2007) are also interested in the use of narratives in the classroom, and they expand on the role of narratives that was laid out by Norris et al. (2005). Metz et al. (2007) focus in on the narrative element of narrative appetite, whose function is to generate a desire to "know what will happen" next. They also are interested in the main purpose of a narrative, which is to help the reader/listener "imagine and feel the experience of others" (Norris et al., 2005, p.545). Metz et al. (2007) are interested in two forms of narrative, stories and story-lines. They claim that stories have a more constrained meaning. It has to have a chronological sequence of events, it must have actions that involve characters, it must have a plot that provides causality between the story's components, and finally, they claim a story is a group of episodes that must be delivered in the same session. Metz et al. (2007) also talk about storylines, they expand on Stinner et al. (2003) definition, discussed earlier in this section, to state that a storyline is a set of chronological episodes that are loosely tied together, and are used to deliver the topic. The characters involved in the episodes, or a thematic topic unify the storyline. The episodes in storylines could themselves be stories capable of standing alone, and therefore delivered over several classes. Apart from defining stories and storylines, Metz et al. (2007) talk about how a story can act as a door-opener. The function of the door-opener is not explanatory, instead they are to make the topic memorable, and to focus the students' attention on the specific point of the story. These stories give the students' "reasons for needing to know" (Norris et al., 2005). These stories are also designed to give the students a mystery or problem to puzzle over, one that encourages them to ask questions. Metz et al. (2007) add to conversion by claiming stories have a more rigid structure that differs from narratives. They also expand the role of stories from just

explanatory to include door-openers. At this point, more researchers have weighed in on the components and definitions of a narrative and/or story. However, what is missing is a way to distinguish between narratives and stories, and how to construct and analyze those stories.

Science story construction. This gap in the research was filled by Stephen Klassen, an innovative science educator, who has studied how stories from the history of physics can promote the learning of and attitudes towards science (2006, 2009, 2010). Klassen (2009) combines Norris et al. (2005) eight narrative elements and two additional narrative elements (effect of the untold and irony) from Kubli (2001) to provide a basis for evaluating stories (See Table 1). Klassen (2009) distinguishes between narratives, which do not need to contain all 10 narrative elements, and stories which do.

Klassen states that text-passages can be analyzed using the 10 narrative elements as a way to find structural deficiencies. He cautions that the narrative elements education. First, the main purpose of a science story is to "improve the teaching and learning of science…" (Klassen, 2009, p. 402). In addition, the reader's response is dependent on the intent of the narrative, if a story is being used as a door-opener, its primary role is "to make the concept being taught more memorable". The type of story Klassen (2009) chose to study is a literary story used as a door-opener. He defines literary story as a brief story, longer than an anecdote (Shrigley & Kobella, 1989) or a vignette (Wandersee, 1990b), that can be judged on both its scientific and literary credentials. The story he chooses to examine has a HOS foundation. Klassen (2009) explains that door-opener stories must provide learners with reasons to engage with the material, a feeling of needing to know

Table 1

Definition of the Ten Narrative Elements

Narrative Element[a]
Event-tokens • Particular occurrences involving particular actors at a particular place and time. • Are chronologically related. • Involve a unified subject and are interconnected. • Later events seen as significant in light of earlier events. • Lead to changes of state.
Narrator • Is the agent relating a narrative (may be in the foreground or background). • Determines the point and purpose of the story to be told. • Selects events and the sequence in which they are told. • Fashions a sequence of events into a significant whole.
Narrative appetite • Desire created in readers and listeners to know what will happen. • Based on a range of possibilities that creates anticipation and suspense.
Past time • Narratives concern the past. • Narrators can manipulate time in relating narratives.
Structure • narratives typically start with imbalances, introduce complications, and end in success or failure • narratives are structured around two independent time sequences- the sequence of plot events and the sequence in which the events are related • Narratives are tied together by satisfying expectations that are established previously.
Agency • Actors cause and experience events in narratives. • Actors are responsible for their actions. • Narratives involve human beings or other moral agents,
Purpose • To help us better understand the natural world and humans' place in it. • To help us imagine and feel the experience of others.
Reader • The reader must interpret the text as a narrative in order to approach it with appropriate expectations and anticipations.
Effect of untold • Missing information increases curiosity.
Irony • The narrative events are in opposition to the reader's expectations.

[a]Adapted from Klassen, 2009.

more. More importantly, the stories should generate questions from the students, as well as leaving a problem or difficulty unresolved, this is the effect of the untold, which creates curiosity. These issues would arise from the events in the story or from the

scientific concepts and processes that the story introduces. He continues by stating that questions are an important part of knowledge acquisition. Another way of generating explanation seeking questions in students is to have the story's situation in opposition to the student's expectations which is caused by the narrative element of irony. Klassen (2009) claims that the presence of this curiosity can be tested in students by having them write down questions that arise from the science story. The potential benefit of this strategy is supported by research that suggests learning is improved when students generate and answer their own questions. Another role of a science story that is they make a point, it is one way students come to understand the story. Student understanding of the point can be tested by having the students write it down after the story is delivered.

Klassen (2009) makes it clear that planning is important before a science story is written, and that the historical components must draw from credible sources and comply with historiographical principles. The story must also relate to the science material being presented and, hopefully to the students' interests. After these items have been attended to, the story must be constructed to include the 10 narrative elements from Norris et al. (2005), and Kubli, (2001) that were mentioned earlier in the paper. The last part of analyzing a science story is to elicit student responses. He suggests having the students write down three questions and the point of the story immediately following its presentation; this format will be unobtrusive if it is designed as an assignment in the lesson. An additional benefit would be to use this assignment as a formative assessment to help focus the reminder of the lesson.

Klassen (2009) applied the analysis protocol on a story he constructed about Louis Slotin, a physicist from Winnipeg who worked on the Manhattan project, who died

from radiation poisoning after an accident in his lab. His motivation for writing this story was the frustration with existing instructional materials on the properties of radioactivity. He felt the original laboratory lesson lacked context, which the topic of radiation protection could provide. He insists that the history used in a science story must be accurate and sensitive to the culture and social mores of the time being portrayed; without creating misunderstandings or suggesting today's science is superior. The story must also be relatable to current students; the goal is to portray scientists as human beings involved in a social enterprise. This humanizing effect can help students become engaged in the story and empathic to the story's characters. Klassen demonstrates how the narrative elements are incorporated into the Slotin story. He condenses the story into just the event-tokens.

> Slotin arrives at work → has doubts about the project → remembers Fermi's warning → demonstrates bomb criticality to Graves and the six other observers → lets the screwdriver slip → is subjected to a massive dose of radiation → reacts quickly to separate the hemispheres → realizes that he have been mortally injured → is attended by Morrison as he dies → is eulogized by Ashlock. (Klassen, 2009, p. 410).

The event-tokens provide the chronology but not motives and choices leading to the events' causative links which creates reader interest in the story. The narrator orders the events sequence and determines the story's point. The Slotin story's narrator is an observer who changes his style from subjective story-teller to commentator. The story's point is to show the importance of radiation protection. The Slotin story uses suspense and foreshadowing to achieve narrative appetite. Foreshadowing was used to predict

Slotin's death and suspense was used during the description of the experiment. The Slotin story relates past events, which is narrative element of past time, Klassen (2009) points out that the story is unique and unrepeatable, which gives it more appeal. The minimal structure of a story is a beginning state which changes due to an event. The Slotin story's structure is framed by the event-token sequence. This sequence itself can be viewed as a string of smaller stories that set the direction for the main story. The Slotin story has agency because it involves characters that make decisions whose consequences they must live with. Klassen (2009) states that the purpose of any story is to help the receiver better understand the world and people's place in it. The Slotin story's purpose, mentioned above, is about radiation protection. The listener/reader should recognize the account as a story and respond to it as a story by wishing to know what comes next. It is this expectation that Klassen (2009) tests by having his students generate questions about the Slotin story. He claims that students demonstrate engagement in the story if they ask how or why questions. The Slotin Story is a short story which cannot include a lot of details and therefore creates gaps in the student's mind that must be filled in. These gaps create the effect of the untold, which increases student engagement. Irony is created in a story when the expectations of the listener are contradicted. Slotin's death creates irony in this story. Klassen (2009) states that irony is an important narrative element, but that it is not as essential as the other narrative elements. He summarizes that educators should adhere to all narrative elements, with irony as the possible exception, so they can become competent science story writers. This is also important for stories to be compared to other forms of prose. These elements only provide structure and creativity is necessary to write a good story. Klassen (2009) used the Slotin story as a door-opener in four laboratory

classes over a period of two years. He told the story while displaying a PowerPoint of relevant photographs, which did not have captions. Following the story, the students were asked to right down the point of the story and up to three questions about the story. The story and the associated assignment took approximately 15 minutes.

At the end of the course, all identifying information was removed from the student responses and they were transcribed into a database. Three assignments were discarded due to inadequate or incoherent responses, so 37 assignments, totaling 104 questions were analyzed. Klassen (2009) began with the assumption that students would ask explanation-seeking questions and that the point of the story would relate to radiation safety. Both Klassen (2009) and a graduate student coded the responses and through consensus two categories emerged. First, all questions were coded by type which included: why, how, what, define, when/who/where, and other questions. Second all questions were placed in domains which included: historical, scientific, ethical, personal and other questions. The analysis showed that about 60% of the questions were why and what types. In the domain categories, 66% were historical and 28% were scientific. Over half (51%) of students did not give a relevant point of the story, of the other 49%, 27% said the dangers of radiation and 22% said the importance of radiation protection were the point.

Klassen (2009) admits the readers may wish to know more about the students' reaction to the story apart from the questions they generated. He explains that the analysis was done after the course was over and therefore no verbal feedback of the students was recorded. He analyzed each response separately, and each student's set of responses was considered as a unit to gain a more holistic view. While most students' set of responses

could not be placed in a single category, one student raised all scientific questions, another student raised personal questions. Klassen (2009) felt the students had varying writing abilities, and his interpretation of the questions could be affected by different skill levels. The discussion on the student responses focuses on two areas, the questions raised by the story and the students' point of the story.

Klassen (2009) reminds the reader that the importance of student-generated questions is that it demonstrates curiosity and that questions are part of theory formation. He states that most of the questions generated where 'when', 'where' and 'who' type questions, which indicated higher order thinking. He feels that the structure of the responses suggest that the students were inexperienced at formulating questions. Most of the student questions were either scientific or historic in nature, there were three scientific questions for every seven historical ones. This ratio may serve as an indicator of the story's specific character. Klassen (2009) recommends that the students have an opportunity to find the answers to both their historically-based and scientifically-based questions through materials or activities. Next the discussion talks about student responses on the point of the story.

Klassen (2009) defines the point of the story as the thesis statement, and as such should possess specific qualities. It should be in a complete statement, have global application, not contain the character's name and express a point of view. For instance, he explains that a theme gives a point of view, but it is not written in a complete statement. A topic simply stated as a fact, or commentaries on the story that do not give a point of view are not thesis statements. Another issue would be giving the moral of the

story, or simply a slogan. The students in this study did not response with a topic or theme, however many slogans and morals were given as the point of view.

Klassen (2009) discusses six expectations for science stories, which together should facilitate learning gains. First, a story should make concepts memorable. The students' responses to the Slotin story indicated they were strongly engaged with the story and this is evidence of the story's memorability. Second, stories help reduce teacher-student distance by making the classroom atmosphere less formal. The Slotin story may have had this effect, but it was out of the scope of this study. Third, a story should illuminate a point. The Slotin story's point "was to illustrate the importance of radiation protection" (Klassen, 2009, p. 418), which would lead to learning about the nature of radioactivity. While some of the students did create a valid thesis statement that aligned with this point; many students failed to meet the criteria outlined above. Fourth, stories should give the students reasons to know the material. The story format provides context to concepts and therefore can illustrate the importance and relevance of these concepts to the students. While not only motivating factor in the study of radiation protection; Klassen (2009) claims the students' responses are strong evidence in support of the story's effectiveness. Fifth, stories should raise student questions and again Klassen (2009) feels the data generated in this study provides evidence that the students asked good questions. Finally, the sixth expectation is that stories should raise explanation-seeking curiosity. Klassen (2009) claims that the student responses to the Slotin story provide evidence that it did raise their curiosity. However, he is unclear how this curiosity relates to new theory formation. Klassen (2009) thinks this connection can be made directly for historical questions For example, in the Slotin story, student

expectations that young, smart scientists will have successful careers was contradicted by Slotin's accident and death. Therefore the students had to revise their expectations to state that young, smart scientists will have successful careers if they follow proper safety standards. Klassen (2009) states this learning sequence is easy to identify, but not very insightful. However for scientific questions it is unclear how student's expectations are contradicted and learning motivated. He thinks the solution to this problem could be in identifying the dominant preconception concerning radiation among the students. The students did not have a sophisticated grasp on the hazardous levels of radiation, and instead they thought any level of radiation is hazardous. Therefore this preconception about the nature of radiation is the students' expectation, and the remaining classroom activities would provide the students with the contradiction that leads to new theory formation.

The Klassen (2009) paper has two main strengths. First, there is a concise argument for the use of narratives in science education and builds on the foundation of Norris et al. (2005). This is important because Klassen (2009) acknowledges the sound theoretical stance of the earlier work and then adds to it by introducing two narrative elements originally discussed in Kubli (2001) and explains the importance of these addition elements from the perspective of learning theory. Second, Klassen (2009) describes a methodology for using these elements to construct and evaluate a science story. He defines the 10 elements and demonstrates how he uses them in the Slotin story. Klassen (2009) divides the story into event-tokens, which helps the reader understand the structure of this important narrative element. He discusses the construction of the story, he relates his strategies for evaluating the effectiveness of the story. The Klassen (2009)

paper advances the research on narratives conceptually by proposing and demonstrating a methodology for constructing science stories. He begins the task of evaluating the effectiveness of the story with student-generated questions, and it is this portion of the process that reveals the limitations to his study.

Klassen (2009) claims that student-generated questions demonstrate that their curiosity was raised and in turn this should pave the way to cognitive dissonance which will promote conceptual change. As admitted above, he was not sure how student curiosity related to conceptual change. One way to see if conceptual change is occurring is to testing the effectiveness of the story on student content gains with a pre and post-test. The study would be stronger if not only content gains were evaluated, but they were done so against a control group. Because none of these occurred in the study, this claim remains in the theoretical domain. Another limitation is the lack of student interviews to clarify their positions on the questions they generated. Interviews may be able to aid in determining a connection between a student's display of curiosity and new theory formation. The interviews could also be tailored to confirm what order of sophistication the student questions were. Klassen (2009) relied on the class of question, such as "how" and "why" versus "who" and "what" questions to determine if the students were displaying higher order thinking. In addition, the motives for the questions they generated could be revealed in an Interview. The overall strength of the study could be increased if interviews were conducted to clarify and verify the student responses.

Purpose of the study. I have designed a study that uses Klassen's (2009) 10 narrative elements to identify and describe the learning outcomes and experiences of participants. I chose this approach because it standardizes construction of stories, thus

giving researchers a way to define and empirically test the efficacy of those stories. To text this approach, one group of participants had an intervention that was not designed with Klassen's (2009) elements in mind, and this is the traditional historical approach, and the second group's intervention was designed to include his elements, and this is the historical story approach. Both approaches are versions of a two-week lesson called the Mystery Phenomenon (MP). This unit uses the history of research on industrial melanism to teach microevolution. The original MP lesson has demonstrated a positive impact on student understanding of nature of science concepts (Rudge et al., 2014), but it has not been evaluated for its ability to impact student learning of natural selection in relation to its narrative elements. Specifically, this study compares student outcomes from two versions of a lesson that utilize some or all of Klassen's ten narrative elements to teach the concepts of micro-evolution. The first version of the two-week lesson employs a traditional historical approach, and the second version employs a historical story approach (Klassen, 2009; Kubli, 2001; Norris et al., 2005). To answer the research questions I conducted a mixed methods study which included a pre- and post-assessment and semi-structured interviews. The research questions are:

RQ1. What differences in learning outcomes are identified in participants' CINS scores in both the traditional and the story approach?

RQ2. What alternative explanations, as identified in the CINS distractors, and the interviews are participants using in both the traditional and story approaches?

RQ3. What are the similarities and differences in participants' experiences, as revealed in the interviews, in both the traditional and the story approach?

RQ4. What do the interviews reveal about the participants' awareness of the story and its narrative elements embedded in the story approach?

CHAPTER III

METHODS

This chapter discusses the research design and methodologies I used to conduct my study. The chapter begins with an overview of the philosophical stance and theoretical framework used to inform this study. The chapter continues with a discussion of the methodological approaches that I used to answer the research questions. This includes the mixed method typology that informed my research design. I discuss the collection, analysis, and interpretation of both the quantitative and qualitative phases of the study. Next, I discuss the process of data integration. I also consider the different aspects of methodological integrity such as the validity and reliability of the quantitative phase, and the trustworthiness and credibility of the qualitative phase. Finally, I examine the limitations of this study.

Methodological Overview

Worldview. I have conducted this study from a pragmatist perspective. The pragmatist worldview has developed from the work of the Peirce, James and Dewey (Cherryholmes, 1992). To find the meaning in an idea, they felt that the useful aspects of the idea must be examined. These pioneers were interested in both the practical consequences and empirical findings derived from the examination of ideas, in part because these things pointed to the next actions that would help uncover the meaning of the phenomenon (Johnson & Onweugbuzie, 2004). Pragmatism has a long history with many facets; one of them is the development of pragmatism as a useful worldview for mixed method research.

Mixed method research draws from both quantitative and qualitative research traditions and the worldviews often associated with these traditions; postpositivism and constructivism respectively, do not easily fit the needs of mixed method research (Feilzer, 2010). Pragmatism has emerged as an alternative to these worldviews as a “philosophical partner” to mixed methods research (Johnson & Onweugbuzie, 2004, p.16). Because pragmatism is not tied to a worldview that endorses specific research methods, but instead is focused on applications and solutions, it allows the individual researcher to choose the combination of methods that work best for the project at hand (Creswell, 2008).

Theoretical framework. This study is informed by the constructivist learning theory as discussed by Driver and Olham (1986). As mentioned above in Chapter One, learning can be considered a process of conceptual change where individuals purposefully and actively construct knowledge in an effort to make sense of the world (Driver & Oldham, 1986). I chose this specific view of the constructivist learning theory because it supports the idea that learning is context dependent and the use of historical stories is a contextual instructional approach.

Research design. I chose to do a mixed method study because both quantitative and qualitative approaches are needed to answer the research questions. Because mixed method researchers have the freedom to tailor their research design to their specific goals there are a great deal of potential research designs available to them. In an effort to organize the wealth of research designs, various researchers have developed typologies (Creswell, 1994, Creswell, 2007; Creswell 2008; Creswell, Plano Clark, & Gutmann, 2003; Leech & Onweugbuzie, 2009; Patton, 1990; Tashakkori & Teddlie; 1998, 2003).

Teddlie and Tashakkori (2009) state typologies in mixed method research are important because they help researchers plan and communicate their research design. However, while existing typologies help researchers plan, there remains several shortcomings to existing typologies. There is no single language in mixed methods research, so researchers must be careful to define their terms. Often typologies are either too complicated or so simple they do not include important criteria (Leech & Onweugbuzie, 2009).

I have selected the typology developed by Leech and Onwuegbuzie (2009). This typology includes three dimensions; (a) the level of mixing, (b) time orientation, and (c) the emphasis of approaches. The level of mixing refers to how the quantitative and qualitative are combined during the research process. There are fully mixed and partially mixed designs. Fully mixed designs combine quantitative and qualitative techniques within one or more stages of the research process. Partially mixed designs separately conduct the quantitative and qualitative phases of the study and combine the data sets at the interpretation stage. The time orientation dimension refers to when the quantitative and qualitative data is collected, relative to each other. In a concurrent design both types of data are collected at approximately the same time, and in sequential designs, one type of data is collected one after the other. The third dimension is the emphasis of approach. In equal status designs both quantitative and qualitative phases have the same weight in addressing the research questions. Leech and Onwuegbuzie (2009) recognize that there are various classification criteria that have been identified by mixed method researchers. These include purpose of study, presence of theoretical framework, time orientation, emphasis of approach and level of mixing. They chose to base their typology on these last

three criteria because they feel these criteria best differentiate between most mixed methods designs. I chose this typology because it describes my research design better than other typologies I considered. For example, the Creswell & Plano typology did not fully describe either my study's purposes or my research design (Creswell et al., 2003). My application of the Leech and Onwuegbuzie typology in represented in Figure 2. The time dimension of my design is concurrent, because neither the quantitative or qualitative data collected was dependent on the analysis of the other set. Both data sets were collected within two weeks of each other;

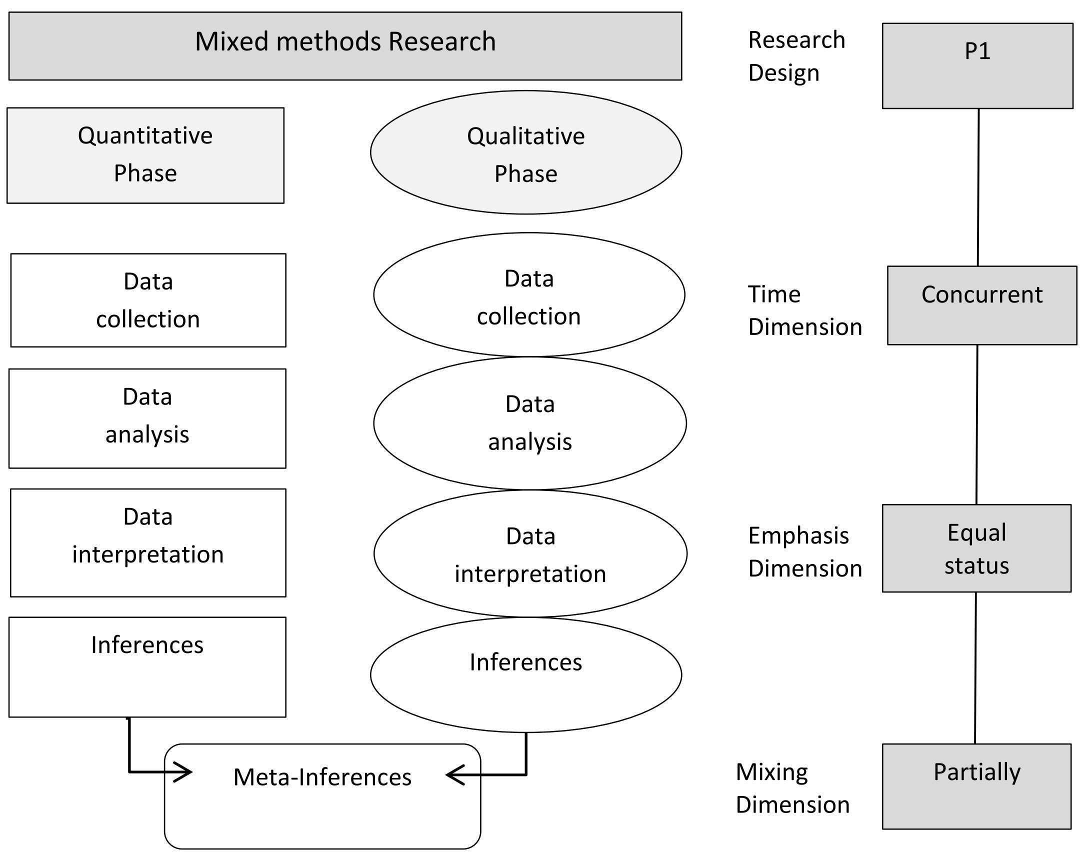


Figure 2. Flowchart of the P1 research design (adapted from Leech & Onwuegbuzie, 2009).

the quantitative data was collected first, followed by the qualitative. The emphasis dimension of my design is equal status, because to fully answer the research questions both quantitative and qualitative data is required. The mixing dimension of my design is partial because I will not be integrating the data sets until both the quantitative and qualitative phases are analyzed. This is a concurrent, equal status, partially mixed design which is also known as a P1design.

Methods

The study was conducted with the approval of the Human Subject Institution Research Board (HSIRB) (Appendix A) at Western Michigan University in the Mallinson Institute's BIOS 1700 course. The study involved six sections of BIOS 1700, three sections per semester for two consecutive semesters. BIOS 1700 is an inquiry-based introductory biology course designed and required for K- 8th pre-service education students.

Participant recruitment. The potential subject pool included the students enrolled in one of the three BIOS 1700 sections in Fall 2013 and Spring 2014. Each BIOS 1700 section holds a maximum of 24 students, 72 students per semester, for a potential maximum of 144 participants. There were two inclusionary criteria for participants. First, they had to be enrolled in one of the BIOS 1700 sections in Fall 2013, and Spring 2014; because the two-week lesson called the Mystery Phenomenon (MP) in BIOS 1700 was being investigated for its potential to influence student learning of natural selection. Second, they had to complete the pre and post-assessment. Study participants were also recruited for a post MP interview. Recruitment for the interview included a verbal

invitation the first day of the unit, followed by an email invitation to any participants who expressed written interest on the consent form. Participants not chosen for the interview were informed by email. The classroom recruitment script and email scripts are in Appendix B. A subset of 30% of the participants was interviewed. There were no exclusionary criteria for subjects.

I gained access to the students by obtaining permission from each section instructor on record through the supervisor of the course. The instructors were responsible for all aspects of the course. The course supervisor had no role in the assignment of grades. I recruited the participants and collected consent forms on the first day of the MP. I conveyed to the students pertinent information about the study including but not limited to: researcher information, reason for the study, process of data collection, participation, short and long-term benefits, potential risks and the consent process). I informed the students that there was extra credit available to those participants who took part in the interview portion of the study. At the end of the MP lesson, students who did not express interest in the interviews and those participants not chosen for the interview were given the opportunity of completing an alternative assignment for extra credit. I provided the students with informed consent forms to review (Appendix A), and reiterated that participation in the study would have no bearing on their course grade, with the exception of the extra credit offered as an incentive for being interviewed. I also provided the students with contact information in case they had any questions pertaining to the study, and then asked interested students to place a signed copy of the consent form in an envelope. The consent forms were not given to the instructors, and are being stored in a secure file cabinet in the course supervisor's office.

Informed consent process. The consent form is an invitation to participate (agreeing to let the data collected on the pre- and post-assessment to be used in the study); it is also an invitation to participate in a semi-structured interview. The consent form has specific details about the study including: participation requirements, data collection (which in addition to the pre- and post-assessment, the demographic questionnaire, and interview recordings, may include all information gathered in normal operations of the course), and potential benefits or risks. Only students who returned a signed consent form participated in the study. In order to protect the participants' privacy, I collected the consent forms and transferred them to the secure location; the course instructors were not aware of who the participants were.

The intervention. The core component of my study is the classroom intervention, and as such it has to satisfy two global objectives. First, as discussed in Chapter One, students often have misconceptions about evolutionary biology, and constructivist learning theories provide strategies to mitigate them. Because this study is informed by constructivist learning theories, it is important to have an intervention that is designed in accordance to constructivism. Second, as discussed in Chapter Two, this study is interested in empirically testing a historically-based science story based on Klassen's (2009) narrative elements; therefore my choice of interventions has to include HOS. In addition to these two global objectives, I chose the MP lesson for reasons specific to this intervention, and these are discussed in the next portion of this section.

The MP lesson was designed with Driver and Oldham's (1986) views of constructivist learning in mind. This means that students must have an opportunity to reflect on their ideas and through a supportive instructional experience, come to the point

of restructuring those ideas (Rudge, 2004; Rudge, et al., 2014; Rudge & Howe, 2009).This constructivist instructional approach is summarized in Table 2. When specifically considering the lesson's NOS learning objectives, the explicit and reflexive approach advocated by Abd-El-Khalick, and Lederman (2000) was used in the unit's design (Rudge et al., 2014). This approach embeds the NOS issues in the instructional design, and the students engage with the NOS issues through the classroom activities and discussions. As a result of the students' own consideration of the issues and classroom discussions, it is hoped they will gain a better understanding of the NOS issues at hand (Rudge & Howe, 2009; Rudge et al., 2014). The MP lesson components and sequence

Table 2

C*onstructivist Teaching Sequence by Driver and Oldham (1986)*

Instructional Stage	Function of Stage
Orientation	Students develop purpose and motivation for the topic.
Elicitation	Students make their ideas explicit.
Restructuring	Students restructure their ideas.
Application	Students use developed ideas in different situations
Review	Students reflect back on their personal process of change.

support the Driver and Oldham instructional approach. For example, on first day of instruction, an opening activity on variation and the following PowerPoint provide the students with a purpose, motivation and context for the lesson. Table 3 lists the components of the MP and aligns them with stages of Driver and Oldham's approach (1986).

The MP lesson is a two-week lesson that is part of a larger evolution unit. This lesson incorporates HOS with reference to research on the phenomenon of industrial melanism to teach students about the NOS and evolution by natural selection. Industrial melanism

refers to a rapid increase in the frequency of dark forms in many moth species that occurred in manufacturing areas affected by large-scale air pollution associated with the industrial revolution in Britain and Continental Europe. As mentioned in Chapter Two, industrial melanism is a well-documented example of evolution by natural selection that is understandable and relatable to students, and therefore an excellent example for teaching (Majerus, 2009; Rudge, 2000). During the two week lesson, the class met four times for two hours and twenty minutes each session. Three of the class sessions

Table 3

The MP's Component Alignment with Driver and Oldham's Teaching Sequence

Day	Components	Teaching Stage
1	Mystery Phenomenon Part 1	
	• Introduction activity: Reflections on Variation (~15 min)	• Orientation
	• Mystery Phenomenon Part 1 PowerPoint (~15 min)	
	• Student Theory Development (~25 min) Student Groups w/handout 1	• Elicitation
	• Movie: Darwin (~30 min)	
	• Class discussion of movie (~20 min)	
	• BugHunt! (~10 min)	
2	Mystery Phenomenon Part 2	
	• PowerPoint & discussions (~120 min) Review of student theories Introduction of 3 scientific theories	
	Theory & experiment discussion Design of Investigations Student Groups w/handout 2	• Restructuring
	Discussion of 3 scientific theories Cooke's theory w/handout 3 Ford's theory w/handout 4 Heslop-Harrison's theory w/handout 5 Film "Evolution in Progress"	• Application
3	Mystery Phenomenon Part 3	
	• PowerPoint & discussions (~60 min) Handout 6: Industrial Melanism Textbook discussion	• Review

explicitly focused on the mystery of industrial melanism in the peppered moth. The remaining class session involved a film on Darwin and evolution by natural selection.

Rationale for use. I have specifically chosen the MP lesson as the intervention in my study because the MP lesson incorporates past scientists' theories and investigations on industrial melanism as a strategy to mitigate misconceptions. The misconceptions targeted by the MP lesson are based on the evolutionary theories of Lamarck and DeVries. Lamarck's Theory of Acquired Inheritance promotes the idea that individuals can adapt to their environment by acquiring new traits within their lifetime, and that these new traits are inherited by the next generation. The population changes gradually over time because the individual organisms are changing in order to adapt to their environment. These ideas are misconceptions because while organisms have the ability to adapt within limits to their environment, there are physiological thresholds in the ability to change. Moreover, changes acquired during an individual's lifetime are not passed down to the offspring. Therefore the changes observed in a population are not due to all the members of that population changing during their lifetimes. DeVries' Mutation Theory proposes that dramatic changes can happen to individual organisms due to genetic mutations. These mutations occur because of a specific agent in the environment, such as a pollutant. The change in the environment can increase the rate of mutations. Population level changes occur because the agents in the environment cause the same mutation in all the members of the population. While mutations can occur to individuals due to environmental factors such as radiation, the increase in mutation rates due to the presence of specific environmental agents has not been supported. The MP lesson discusses these theories along with Darwin's theory of evolution by natural selection, in conjunction with

scientists who investigated the change of the *Biston betularia* population in accordance to one of these theories.

The following summary explains how the MP lesson uses the history of research on industrial melanism to reveal and alleviate student misconceptions (adapted from Rudge et al., 2014). The first day of the MP lesson introduces the students to a mystery about the peppered moth, *Biston betularia* which is common throughout Britain and Continental Europe. The goal for this session is to elicit the students' ideas, including misconceptions, of how this mystery could have occurred. The light, speckled form of the moth was first described by the naturalist Moses Harris in 1766, and at this point no other form of the moth was known to exist. The students view a photograph of the moth resting on a lichened-covered tree trunk, which demonstrates how well the moth is camouflaged against the background, suggesting there is an adaptive advantage. The discovery of a dark form of the moth ca.1848 near Manchester, England is now shared with the class. This discovery encourages naturalists to search for more specimens. Next the students look at a series of survey maps spanning 100 years that reveal continually increasing sightings of the dark form, and the instructor points out that the dark moth sightings seem to be only in certain areas. Students are also informed that not only is the range of the dark moth increasing; in some areas where is has been previously unknown, it is now very common.

Students are now asked to think about what else could have been happening in Britain and Continental Europe from 1850 to 1950. At least one student always responses that the Industrial Revolution was occurring during this time period. The Industrial Revolution represents the rapid increase of coal, coke and oil- powered manufacturing

centers that resulted in large scale air pollution. The students are able to see the change through time (1730, 1860 and 1954) in the landscape near Manchester by viewing three images of approximately the same area. At this point the instructor displays a map that documents the relative frequencies of light and dark forms throughout Britain, and asks if there is a noticeable pattern. Students do notice that the dark form is most common in the vicinity of large cities such as London, Manchester and Edinburgh, and the instructor confirms that these cities were the manufacturing centers during this time period.

The instructor asks the students to explicitly discuss why they think the dark form is becoming more common near the manufacturing centers, and to provide reasons why their ideas are plausible. The instructors' experience teaching the MP lesson shows that students provide one of three possible explanations to the mystery. Some students understand that natural selection is the correct explanation, but they may have trouble verbalizing it. The rest of the students offer explanations that are considered misconceptions, such as individual moths will change colors if needed (Lamarck's theory), or that the ingestion of specific pollutants by the caterpillars will directly cause mutations (Devries' theory).

The second day of the MP lesson starts with a recap of the previous day, which includes the three explanations of the mystery phenomenon generated by the students. The goal of this session is to validate the students' ideas, which is accomplished by the instructor who reveals that the students' explanations were each proposed by a past scientist who did research on the mystery phenomenon. The instructor also shares that these scientists (E.B. Ford, Nicholas Cook and James Heslop-Harrison) developed their hypotheses in context of the theoretical frameworks of Darwin's Theory of Evolution by

Natural Selection, Lamarck's Theory of Acquired Inheritance and DeVries' Mutation Theory.

A classroom discussion about what theories are in general follows this introduction. The goal of this *explicit reflective* discussion (c.f. Abd-El-Khalick and Lederman, 2000) is to have students share their genuinely held ideas about theories without judgment. The students are asked to talk about other theories they have encountered in other science classes, and what about these examples makes them a theory. The class is allowed to reach some consensus on the nature of theories, and then they are asked if and how scientists choose between different theories. The students generally know that scientists conduct investigations, including experiments to test competing theories. The Instructor asks the students if scientists have other ways of choosing between different theories, anticipating that some students may consider the plausibility of the theories, and consistency of the theories with other more accepted theories. The second day ends with the students generating ideas about how each of the proposed hypotheses may be tested by observations and experiments.

The third day also begins with a recap of what has been discussed about the mystery phenomenon up to this point. The goal of this session to have the students discuss how each of the three explanations might be tested with reference to the results of similar tests conducted by past scientists. To accomplish this, students are asked to work through summaries of past investigations that represent each proposed explanation, and to do so from the perspective of both proponents and critics. The instructor emphasizes the role of observations and experiments in the testing of hypotheses, and this provides an opening to start a discussion on the nature of experiments. This *explicit reflective*

discussion is designed to encourage the students to share their ideas about experiments without judgment. As with the discussion about theories, students are asked to talk about experiments encountered in other science courses, and to discuss what makes their example an experiment. The students' examples gives the instructor an opportunity to focus on the restricted use of the term "experiment" in biology.

> i.e. in the life sciences, 'experiment' is used specifically to refer to systematic study in which a system is perturbed with reference to an independent variable and the effect of the perturbation is observed with reference to a dependent variable. (Rudge, et al., 2014 p. 1886).

When the class comes to consensus on what an experiments are, the students are asked if experiments are always necessary, and the focus of the discussion is on the students' examples. The instructor encourages the students to think about historical sciences, such as geology and evolution where experimentation is not always possible. After the discussion the class watches *Evolution in Progress*, a film that seems to establish that the mystery phenomenon should be explained in terms of natural selection (Kettlewell, 1961). In the last portion of this session the instructor shares details about the mystery phenomenon that reveals it is more complicated than textbooks portray. This leads to the final discussion about how the students, most of who are future elementary teachers, would help their students understand how misleading textbooks accounts can be in terms of the process and nature of science. This summary highlights two important characteristics of the MP lesson. First, the students are exposed to the theories and investigations from the history of research on industrial melanism through visual images, activities and discussions, and these components focus on the possible explanations for

the changes in the *Biston betularia* population. Second, some of the possible explanations are misconceptions that both the students generated, and were used by past scientists as legitimate hypotheses. The lesson not only explicitly talks about the scientifically appropriate explanation of natural selection, it discusses past theories related to currently held misconceptions, in this way the correct explanation is reinforced while the misconceptions are minimized. The inquiry format of the lesson has the students work through possible solutions in order to make an evidence based decision of the best explanation.

While I chose to use the MP lesson for the above reasons, there were additional benefits to using this lesson. First, the MP lesson has been used for several years, and it has demonstrated success in the teaching of NOS aspects (Rudge, et al., 2014) which I feel makes it a promising candidate for further research on the unit's content. Second, the same instructors taught equivalent course sections in both the fall and spring semesters. The three instructors have also taught the course in past semesters and were comfortable with the material. Finally, the structure of the unit allowed for modification of just one set of components, the scripts and the accompanying PowerPoints, which enabled me to test the effectiveness of the modifications.

***MP modifications*.** The original MP lesson is historically based, but it was not designed with any of Klassen's narrative elements in mind, therefore if the original MP lesson has any structural similarities to a story it is because historical accounts naturally include some narrative elements. It is precisely the fact that the original MP lesson was not designed as a story that it could be used as the intervention in the control group. The original version of the lesson delivers the historical account of industrial melanism

through a combination of three PowerPoints and accompanying scripts. While the original scripts did contain most of the narrative elements, the problem was that they were not tied together in a cohesive unified whole, and this gave the lesson the feeling that each day was a separate account. For example, event-tokens are supposed to tie previous events to later ones through a causal relationship and therefore provide the structure for a story. If there is not an adequate sequence of event-tokens in an account, the ones that are present may be isolated from the rest of the account. This is the case in the original version of the MP lesson, there were isolated event-tokens sprinkled throughout the three scripts, but they were not tied to each other. Another problem was that while there were many examples of agency in the original version, these occurrences were generally not related to any event-tokens. While the original version was appropriate for the control group, the limitations with reference to Klassen's narrative elements meant that it needed modifications to make it suitable for the treatment group. The treatment's group intervention had to be designed to specifically contain Klassen's narrative elements in such a fashion that the participants would hopefully demonstrate learning gains and report positive experiences with regards to the story. Table 4 summarizes the limitations of and modifications in reference to Klassen's (2009) elements in both versions. In summary, the purpose of this study was to identify and describe the students' learning outcomes and experiences in a historical story that purposefully contains Klassen's narrative elements. The four research questions designed to achieve this purpose needed both the original MP lesson which was not designed with Klassen's elements in mind, and one specifically designed to include these elements.

Table 4

Limitations in Specific Points of the Narrative Elements in the Original MP Scripts and the Changes Made to the Modified MP Scripts.

Specific Points of the 10 Narrative Elements	Limitations of Original MP Scripts	Changes in the Modified MP Scripts
Event tokens		
• particular agents	• anonymous agents "Moth collectors" (S1,S9)	• named agents added "Collectors such as Tutt" (S1, s9)
• related by unified subject	• event tokens isolated "EB Ford built off Tutt theory (S2, S21)	• event-tokens added to connect isolated events "EB Ford collected butterflies and moths...became very interested in different varieties. "He decided to study biology…as a result…" "He (Ford) built off of Tutt's theory…" (S2, S17)
Narrator		• present linked to past "In Britain, people are avid collectors…" (S1, s3) "Butterfly and moth collecting in Britain has been going on for centuries…" (S1, s4)
• selects events & sequence	• does not tie past to present "The moth was first characterized in 1766…" (S1, s5) "In 1848 a different dark colored form was discovered…" (S1, s7)	
• creates significance & meaning	• lacks relevance to students' lives no information about scientists as educators	• scientist as educator thread added "He (Tutt) was a school-master who loved…and teaching…" (S1, s9) "He (Heslop Harrison) was an energetic, highly- praised high-school science teacher…" (S2, s19)
Narrative appetite		• scientists' background added "Ford, like Tutt a ½ century before, passed his interest…" (S2, s17) "He earned his doctorate and moved on…" (Harrison: S2, S19) "At the age of ten he bought
• desire for what happens next	• scientists not humanized no background information to create interest	
Past time		• account starts in present and returns to the past Present day moth collecting is tied to past collectors. (S1, s2-8)
• time manipulation	• account only related in chronological order	

Table 4- continued

Specific Points of the 10 Narrative Elements	Limitations of Original MP Scripts	Changes in the Modified MP Scripts
Structure • Tied together by satisfying previously established	• MP solutions not related through event-tokens	• Added event-tokens to tie scientists and solutions to each other. "Kettlewell's work is compelling." "However, Kettlewell's conclusions were not accepted by all scientists." (S3, s16) "Majerus critiqued Kettlewell's work, but he was supportive of bird predation…" "He therefore conducted experiments…" (S3, s18)
Agency • agents cause and experience events in narratives	• agents not attached to event- tokens "He was convinced that lead salts in the soot might have mutagenic properties." (Harrison: S2, s23)	• more event-tokens built around existing agents "… he was a loner and did a great deal of work from home. He had an uncompromising personality and he was always convinced of the correctness of his theories and views. " "He was convinced that lead salts in the soot might have mutagenic properties." (Harrison: S2, s19)
Purpose • To help us imagine and feel the experience of others	• Lack of event-tokens and background information limits ability to experience scientists' lives.	• Added more event-tokens and background information on the scientists to humanize account. "Tutt was a high energy, amateur entomologist…" "He could be seen in the field with his pockets stuffed with tins full of butterflies and moths;" (S1, s9).
Reader • must recognize account as narrative	• account not written as a narrative	• modified account includes all narrative elements
Effect of untold • missing information	• lesson designed as a mystery	• no modifications
Irony • events are in opposition to the readers expectations	• Introduction of 3rd moth	• Introduction of 3rd moth • Scientists' lives do not proceed as expected. "…may have contributed to his early death in 1911." (Tutt: S1, s9) "…not able to publish his results…died unexpectedly in 2009 at the age of 53." (Majerus: S3, s19)

Therefore the limitations found in the original version needed to be addressed in the modified version.

To address the limitations of the original MP lesson, I needed to make changes to two MP components, the original MP scripts (Appendix C), and the accompanying PowerPoints (Appendix D). It is these two components together that make up the historical account of industrial melanism. There are two foci of the modifications. The first is to create a historical story of industrial melanism that contains Klassen's ten narrative elements while maintaining the objectives of the original historical account. These narrative elements are discussed in Chapter Two. Klassen also reminds us that a story is more than its narrative elements, they must be put together in a way that creates a coherent and interesting whole (2009). Stories add context and humanize the content, which makes it more relevant and interesting, and in turn may motivate students to be more engaged with the material (Klassen, 2009; Metz et al., 2007). Therefore, the second focus of the modifications was to make the original account more relevant and relatable to the students. To accomplish this I made the story more current and enhanced the human element of the researchers. Examples of these changes are below.

The modification of the scripts and PowerPoints occurred in three stages. In the first stage, I examined the original scripts and PowerPoints. I reviewed the scripts to determine the type and location of each narrative element. I also took note of where additional narrative elements and other pertinent information could be inserted. I reviewed the PowerPoints to gain an understanding of how they supported the script, and to determine what could be changed to enhance this support. I also took note of any additional information provided in the PowerPoints that added to the historical account

given in the scripts. I also had to determine what narrative elements were delivered by each instructor. To achieve this, each day of the MP for all course sections were videotaped. I reviewed the videos and recorded the type and location of each narrative element. The videos were also helpful in finding any areas in the original account where the instructor had difficulties in the delivery or flow of the lesson. For example, the classroom videos revealed that the instructors had difficulty explaining peoples' interest in moth collecting 150 years ago. My review of the scripts and PowerPoints for narrative elements and other information enabled me to formulate a modification plan.

The second stage involved the planning and implementation of the changes to the scripts (Appendix C) and PowerPoints (Appendix D). I revised the scripts first because the modified scripts served as a template for restructuring the original PowerPoints. The modified scripts had to contain all 10 of Klassen's narrative elements in a way that makes the story more relevant and relatable to the students. In order to make the story more relevant I wished to bring the MP research story into the present day, and therefore I added Majerus' research to the scripts (Cook, Grant, Saccheri, & Mallet, 2012). Second, to make the story more relatable I wished to humanize the naturalists and scientists involved with the peppered moth, so I added personal and professional information to the scripts (Salmon, 2000). I made these changes with the 10 narrative elements in mind, and Table 4 gives a synopsis of what narrative points were missing in the original scripts (See Table 1 in Chapter 2 for a more detailed account of Klassen's narrative points).

After the scripts were rewritten, I modified the original PowerPoints. There were several reasons for these changes. First, The PowerPoints had to align with the new scripts. For example, in modified PowerPoint 1, I had to update the pictures in slides 3-4

so they would align with the text in script 1. Second, I added new graphics to the PowerPoints to make them more user-friendly. For example, I added timeline to the bottom of the slides in all three PowerPoints so the students could easily follow the story over time. Third, I updated the PowerPoints' background colors and text fonts for a more visually appealing affect. Overall, I revised both the scripts and PowerPoints to incorporate the story and its supporting information in a manner that maintains a smooth flow through all three days of the MP lesson.

In the third stage of the modification process, I demonstrated the new materials and gathered feedback for final changes. Before the beginning of the MP lesson in Spring 2013 I sent the modified scripts and PowerPoints to the course's supervisor and the three instructors. This gave them an opportunity to review the materials before I delivered them in the classroom setting. A week before the start of the MP lesson, I reviewed the scripts and PowerPoints for the course supervisor, the three instructors, and other graduate students. They gave feedback on both the scripts and PowerPoints, if there was agreement, the changes were made. The final version of the scripts was given to the instructors and the PowerPoints were installed on the classroom computer.

Data collection. This research study was conducted over two semesters and both quantitative and qualitative data will be collected in both semesters. In keeping with the P1 research design described earlier, the time dimension for data collection is concurrent. Therefore, I collected both the quantitative and qualitative data within approximately the same time. The quantitative data was collected on the first day of the MP lesson and the day after the completion of the lesson, and the qualitative data was collected within a week after the conclusion of the lesson. While the interviews followed the administration

of the pre and post assessments, this was because the interview questions pertained to the MP lesson, and therefore had to take place after the lesson, but the interview questions were not dependent on the participants' answers in the assessments, and therefore the data collection was not sequential in design.

Quantitative data. I have two different data sources to assess the participants' understanding of natural selection. The principle source of data is the Conceptual Inventory of Natural Selection 2013 high school/college version (CINS), which I used as the pre and post assessment of the MP lesson. The CINS is a forced answer instrument with one correct answer and 3 distractors. It contains 20 questions and covers ten concepts about or related to natural selection (Table 5).

The original CINS was developed to be used in the classroom to assess student understanding of the process of natural selection and the foundational genetic and ecological concepts needed to understand natural selection as an explanatory theory. The CINS uses actual scenarios studied by scientists which provide a realistic quality to the test items. It has 20 forced choice questions, two for each concept. The distractors used in the CINS are common alternative conceptions and therefore the inventory not only measures student knowledge, it identifies problem areas that the instructor can address (Anderson, Fisher & Norman, 2002). The current CINS is a modified version of the 2004 version of the original CINS. It keeps the structure and overall feel of the previous versions, but provides high school and college teachers with a more organized and readable instrument (Evans & Anderson, 2013).

I chose to use the CINS instrument for several reasons. First, the instrument was designed based on constructivist and socio-constructivist learning theories (Anderson et al., 2002). This is important because my study in general and both versions of the intervention were designed using constructivist learning theories. Second, part of RQ1 is concerned with explanatory coherence, which is how consistent students are in answering different questions about the same concept (Kampourakis, 2009). The CINS has two questions per concept, so changes in explanatory coherence can be measured using this instrument. Third, when considering RQ2, the CINS is useful because the distractors are common alternative conceptions. Anderson et al. (2002) classified each distractor according to these alternative conceptions. For example, some of the distractors appeal to the wants or needs of the individual to change in order to survive. These distractors suggest intentional or teleological thinking. Other distractors talk about learned behaviors being inherited or traits changing do to use or disuse, and being passed on to offspring; suggesting Lamarckian ideas of evolution. The type of distractors picked by the participants may provide insight to what type of explanations they give. The data analysis section of this chapter discusses how participant distractor choice will be analyzed. Fourth, all versions of the CINS have been extensively field tested and validated. The Original CINS was tested for face validity, internal validity and reliability. Face validity was verified by independent content experts. Internal validity was measured by principle components analysis and it was found that the CINS's measurement structure was validated by the analysis. The reliability of the CINS was measured by the Kuder-Richardson 20 and the results were found to be acceptable. Readability of the question stems was also considered, and was found to be realistic for beginning college students.

The current version of the CINS, whose main changes involved improved readability and rearranging the question numbering system so the instrument can easily be broken into a pre and posttest. The CINS was field tested on both high school and college students. The results support that this latest version is effective in assessing concept understanding (Anderson et al., 2002; Anderson & Evans, 2013). After comparing the 2004 version and the CINS, I feel the changes to the arrangement and readability are significant enough to

Table 5

The Concepts and Definitions of the CINS 2013 High School/College Version

Concept	Definition
Biotic potential	All species have such great potential fertility that their population size would increase exponentially if all individuals that are born would again reproduce successfully.
Stable populations	Except for minor annual fluctuations and occasional major fluctuations. Populations normally display stability.
Limited resources	Natural resources are limited. In a stable environment, they remain relatively constant.
Limited survival	More individuals are produced that can be supported by available resources, but population size remains stable; therefore a struggle for existence among the individuals of a population results in the survival of only a part of the progeny of each generation.
Variation	No two individuals are exactly the same; rather, every population shows enormous variability.
Origin of variation	New variation appears randomly through mutation and sexual reproduction. [a]
Variation inherited	Much of the new variation is heritable.
Differential survival	Survival in the struggle for existence is <u>not random</u>, but depends in part on the hereditary constitution of the surviving individuals. This unequal survival constitutes a process of natural selection.
Change in population/ origin of species	Over the generations the process of natural selection will lead to a continuing gradual change of populations, that is, to evolution and to the production of new species.

Note[a.] Table adapted from the Conceptual Inventory of Natural Selection 2013 High School/College Version developed by Anderson and Evans (2013)
Note[b] Concept included in CINS because it is essential for the process of natural selection, however it occurs prior to natural selection.

the CINS, I feel the changes to the arrangement and readability are significant enough to warrant using the latest version. Finally, I am using the CINS because it can be administered in a fairly short amount of time, about 20-25 minutes, which will keep the disruption to the class to a minimum.

In addition to the CINS data, I collected results for seven questions in Group G of the BIOS 1700 Final Exam. These seven questions explicitly refer to the MP. Five of the questions are from the pilot version of the CINS, and two were generated from the BIOS 1700 course supervisor. These questions were evaluated because they are a potential source of data for the participants' learning outcomes.

Qualitative data. I conducted semi-structured interviews (Appendix E) with a subset of the participants. I interviewed 15 of 41 participants for Fall 2013 semester and 14 of 46 participants for Spring 2014 semester. The purpose of the interviews is to help answer research questions RQ3 and RQ4 which are concerned with participant understanding of MP content and participant experiences with the MP lesson. The interviews were about 30 minutes in length and took place at a mutually agreed upon time. The interviews were conducted in the Mallinson Institute for Science Education library located in Wood Hall on Western Michigan University's campus. The interviews were audio-recorded. The interviews were conducted within a week after the completion of the MP lesson. This time frame was important because the participants must have experienced the MP, but not have completed the Final Exam which includes specific questions on the unit and could be a confounding factor.

Data analysis. As with the data collection phase of my study, the data analysis occurred concurrently, and therefore I analyzed the quantitative and qualitative data independent of one another and approximately at the same time. For the quantitative portion of the study I employed a quasi-experimental design with a nonequivalent control group. The CINS was used as a pre and post-assessment at the beginning and end of the MP lesson. A summary of the CINS analysis procedures relating to the research questions are in Table 6. The qualitative portion of the study involved conducting semi-structured interviews with a sub-sample of participants in both the control and treatment groups. A summary of the interview analysis relating to the research questions is in Table 6.

Quantitative data. I have used both descriptive and inferential statistics to analysis the quantitative data. Before I could evaluate the data, it needed to be organized in data files that would aid in the management and manipulation of that data. I used Excel files to manage my data. While I was creating the Excel files, I was also going through the process of cleaning the data, which is checking for data entry errors (Shutt, 2009). After the data files were created, I applied descriptive statistics to find the mean, variance and standard deviation for the CINS scores, CINS distractor scores, CINS explanatory coherence scores, Exam 4 scores, and participant GPAs. In addition, I evaluated gains and normalized gains for the CINS scores for each semester. These descriptive statistics define the distribution of the sample. To aid in the interpretation process, when appropriate, I graphically represented the distribution data, and this is discussed in Chapter 4. I also used cross-tabulations to determine the relationship between variables. I applied McNemar’s test to evaluate dependent data pairs, and chi-square analysis for the independent data.

Table 6

Analysis Procedures for the CINS Instrument

Research question	Analysis procedures
Q1 Differences in Participant scores & Explanatory coherence	• A paired t-test determines if there are significant differences within both the control and treatment group. • Change can be considered statistically significant if the p-value < 0.05 • An independent t-test determines if there are significant differences between the control and treatment groups • Change can be considered statistically significant if the p-value < 0.05 • McNemar's test determines what differences exist in for each CINS question in both the control and treatment group. • It tests for significant changes in correlated proportions. • It reveals what individual questions the participants significantly improved on from pre to post-assessment (Adedokun & Burgess, 2012) • Normalized gains (ratio of actual average gains to maximum possible average gains) of the CINS scores for both groups will determine instructional effectiveness of the control and treatment groups (Hake, 1998). • The effect size of the CINS scores for both groups determine the effectiveness of the intervention of the control and treatment groups (Coe, 2002). • Effect size is will be calculated using Cohen's d.
Q2 Alternative conceptions	• The CINS distractors were coded using a priori codes for common alternative explanations. • The summation of distractor choices for both groups will determine what alternative explanations the participants are using.

Descriptive statistics are not only useful to illustrate distributions; they are needed for the application inferential statistics. For example, I used the gains to perform paired t-tests on the CINS scores, the CINS distractor scores, and the CINS explanatory coherence scores for each semester. Also, I used the mean to perform independent t-tests to the all the scores between semesters. I interpreted the results of both the descriptive and inferential statistics to answer research questions Q1a and Q1b, which is reported in Chapter four.

For the CINS scores, I calculated the values for all twenty questions and then for the ten targeted questions. My rationale for calculating the values both ways is connected

to the ten natural section concepts covered in the instrument (Table 5). The CINS was designed to encompass the logic of the theory of natural selection, therefore if the students choose all correct answers they have a good understanding of how populations are affected by natural selection (Anderson et al., 2002). Therefore to ascertain how well the participants understood the theory of natural selection, all questions needed to be considered. Four of the ten concepts were taught directly before the MP lesson, five were covered in the MP lesson and the final concept was taught after the MP lesson, therefore it is appropriate to evaluate the participants on all 20 questions because they received explicit instruction in 9 of 10 concepts by the end MP lesson. Two objections to this rationale could be voiced. First, because the MP lesson only explicitly covered five concepts, those are the only ones that should be considered. But as stated above, the instrument was designed in part to determine the how students understood the theory of natural selection overall, and to do so, all questions need to be considered. This objection could be a potential concern if the participants did not have instruction on four of the five concepts not covered in the MP lesson, however because they did, this objection is not concern for this study. Second, there was one concept not extensively covered before or during the MP lesson, it was mentioned in a movie shown during the MP lesson, and an argument could be made that this concept, should not be considered in the results. This objection ignores the fact that students come into the classroom with prior knowledge, and the pre and post-test design will determine if student understanding of this concept was affected by the MP lesson. While the CINS instrument is valuable for determining the overall understanding of natural selection, it is a criterion based test which means student understanding of individual concepts can be determined, and therefore since the

MP lesson was focused on five concepts, I wanted to determine if there were differences in the scores of these concepts. The five concepts explicitly covered in the MP lesson are: (a) variation, (b) the origin of variation, (c) certain variations are inherited, (d) differential survival and, (e) change in populations. While looking at the five concepts can help determine how effective both versions of the MP lesson were for specific concepts, the interpretation of the results will focus on the 20 question format because this gives a more holistic view on how each version influenced the participants' understanding of the theory of natural selection.

Qualitative data**.** As mentioned earlier, I interviewed a subset of the participants from each semester. In Fall 2013, the interviews consisted of seven questions focusing on the context and content of the MP lesson, and the participants' experiences of the lesson. The interviews for Spring 2014 had two additional questions aimed at the modified PowerPoints and the concept of a story. The qualitative phase has both an exploratory and explanatory portion. (Marshall & Rossman, 2011). The exploratory portion wanted to discover what meaning the participants made of their experiences with the MP lesson, and the explanatory portion wanted to see if the participants' responses would add insights to their answers on the CINS instrument (Table 7).

I began the process of transforming the data by transcribing all of the interviews. I transcribed each participant's interview in its entirety, using Word as the processing software. For each semester, I copied all the transcripts to a separate Excel file. To facilitate the coding process I also set up Excel files for each semester by individual question. The Excel program has functions, such as conditional cell formatting, that aid in discovery of relationships between data points (Meyer & Avery, 2009). After the

transcriptions were done, I moved on to coding the interviews. According to Saldaña, coding can be divided first-cycle and second-cycle coding methods. First-cycle methods are preliminary approaches to organizing and getting a feel for the data. Second-cycle methods help the researcher make meaning of the data. This is achieved through a variety of processes; such as classifying, prioritizing, integrating and synthesizing the data. While coding methods can be classified as first or second-cycle,

Table 7

Analysis Procedures for the Semi-Structured Interviews

Research question	Analysis procedures
RQ2 cont. Participant explanations	• Emergent coding was applied to discover what explanations were given by the control and treatment group. The participants' responses were cross-analyzed to explore similarities and differences within and between groups.
Consistency of explanations	• The interview explanations of MP content were coded using a priori codes in both the control and treatment group. • The a priori codes for the interview explanations and CINS answers will be cross-analyzed to determine if there was consistency in both the control and treatment group. • The coded interview explanations and coded CINS distractors will be cross-analyzed to determine the consistency in the use of alternative explanations in both the control and treatment group.
RQ3 Participant experiences	• Emergent coding was used to identify and describe participants' experiences in both the control and treatment group. • The coding results within both the control and treatment groups will be compared to discover the similarities and differences in the participants' experiences. • The results for the control and treatment group will be cross-analyzed to discover the similarities and differences between the groups.
Q4 Story elements	• Emergent coding was used to discover if the participants' experienced the treatment intervention as a story and to identify what components the participants consider part of a story.

the movement from a first-cycle method to a second-cycle method is not a linear process or even required. The research project will guide what methods are used and how many coding cycles are needed (2009). Marshall and Rossman discuss how coding can be

ranked on a continuum that ranges from prefigured to emergent. In prefigured coding, categories are determined in advance of engaging with the text, while in emergent coding, the researcher generates codes through immersion with the text (2011).

My data analysis process involved two distinct rounds of coding that corresponded with the exploratory and explanatory portions of the research. As mentioned above, in the first round I was interested in what insights the data would reveal about the participants' explanations and experiences with the MP lesson. Therefore to analysis the data in this first round, I used emergent coding in an immersion analytical style (Marshall & Rossman, 2011) and a combination of first and second-cycle coding methods (Saldaña, 2009). I chose to perform the emergent coding in the first round so bias from second round, which involved a priori codes, would not be introduced. Emergent coding was performed on all interview questions for both semesters, and this first round of coding helped answer research questions RQ2 – RQ4. In the second round of coding, I was interested in comparing the participants' responses to specific interview questions to their answers on the CINS instrument. Therefore, I used a template analytical style with a priori codes for this second round of coding (Marshall & Rossman, 2011). The second round of coding was performed on interview questions IQ2 and IQ3 for both semesters, and this coding helped answer the second portion of research question RQ2.

In the first round of coding, I began the coding process with the first-cycle method of attribute coding. Attribute coding involves the organization of the specific information related to the collection of the data. For example, on each interview transcript and corresponding Excel file, I included the participant's initials, class session,

date and time of interview and the length of the interview. Attribute coding helps with proper data management and assists the researcher in timely retrieval of participant information. It also provides context to the coding process which is needed for data analysis and interpretation (Saldaña, 2009). Once I completed attribute coding, I moved on to another first-cycle method called initial coding. I used initial coding to closely examine the data for patterns from which themes and categories emerge. Initial coding is an open-ended strategy, whose purpose is to help the researcher remain aware of the possible directions the data may take. Since coding is a highly iterative process, I continued the cycle of coding until no more unique codes were being generated. At this point I made the decision that the coding process had not led to sufficient understanding of the data, so I employed second-cycle method of focused coding to further reduce my data and to help generate categories (Saldaña, 2009).

A priori codes. As mentioned earlier, to answer the second portion of RQ2, I needed to develop a priori codes so that the CINS instrument choices and the interviewee responses could be compared. The purpose of this research question was to compare the interviewees' answers on the CINS questions to their interview responses on variation and natural selection. Once the a priori codes were developed, I first applied them to the interviewees' answers to the CINS instrument and second, after emergent coding was completed, to the interviewees' responses to IQ2.

Development of the a priori codes. There were several tasks involved in the creation of the a priori codes. As discussed above, Anderson et al. (2002) based the distractors for each question on common alternative conceptions, and it was these statements that I used to develop my a priori codes. My first task was to match the

questions in the current CINS version to the original CINS. The original version had been renumbered to have easily discernible pairs of questions that covered one concept. For example in the original version the two questions that covered variation were #9 and #16, and on the current version they are #5 and #15. This alignment was important because the alternative conceptions were described only for the original version. Therefore, to ascertain if our coding scheme agreed with Anderson et al. (2002) alternative conceptions, I needed to match the original CINS questions with the current questions. Second, I compared the current version's answer choices to original version to make sure they were aligned. Finally, I aligned the alternative conceptions in Anderson et. al (2002) to the current distractors. For examples of two concepts that were aligned, see Table 8.

Once the alignment of the question numbers, answer choices and alternative conceptions to distractors was complete, I was able to develop the a priori codes. First, I placed alternative statements into groups of similar concepts, and from there developed tentative codes that all the statements fit into. Next, I discussed the tentative codes with my second coder and we clarified any ambiguities. We each separately coded questions 5-19, which were the 10 questions targeted in the MP lesson. At a second meeting we went over the coding and discussed discrepancies between the coded questions. At this point we were able to reach consensus on 93% of the coded answer choices, and these aligned with Anderson et al., 2002. However, we did not agree on the a priori codes for three answer choices. To resolve this, I gave preference to the a priori code that aligned with Anderson et al., 2002. There were two problematic answer choices on the concept of change in population, questions 9c and 19c. Both coders assigned the a priori code ACQ for each answer choice, which is about the inheritance of variations (Table 9). This code

does not align with the concept of change in population. The after careful review of the answer choices of 9c and 19c, and the Anderson et al., 2000 alternative conception attached to those choices; I was not able to change the a priori codes we assigned. Therefore these answer choices were not used in any final analysis for RQ2. This is discussed further in Chapter Four.

Application of a priori codes. Once the priori codes were developed they were applied to the interviewees CINS responses. Once emergent coding was complete, the interview responses were coded again using the a priori codes. The a priori codes assigned to the interviewees' CINS answers were compared to the codes assigned to the interview responses, and this comparison was evaluated to answer question RQ2. The results are discussed in Chapter Four.

Data interpretation. Data interpretation is the process of bringing meaning to the data (Marshall & Rossman, 2011). This is important for two reasons. First, the interpreted data is used to answer the research questions, and second, inferences generated from the data interpretation may also provide new understandings and explanations of the phenomenon under study (Tashakkori & Teddlie, 2008). I interpreted the quantitative data after analysis, which is consistent with my P1 research design (Figure 3) On the other hand, data interpretation of qualitative data is part of analysis, interpretation of the data occurs throughout the entire process. (Marshall & Rossman, 2011). At the end of the interpretive process, I used the inferences derived from the data to answer the research questions, which are discussed in Chapter Four.

Data integration. I integrated the findings from the quantitative and qualitative portions of my research after all the data was analyzed separately. Integration is the process of combining both data sets. I have two foci for data integration. First, both data sets were compared in an effort to answer research questions RQ2. Second, I examined the data sets for similarities and differences in an effort to identify any emerging meta-inferences (Tashakkori & Teddlie, 2008).

Methodological integrity. In mixed method research, both the quantitative and qualitative portions of the study must be evaluated by their own set of quality standards. In addition, the meta-inferences made during the data integration process must be evaluated for quality. I considered four areas of integrity; instructional fidelity, the validity and reliability of the quantitative data, the trustworthiness of the qualitative data, and the quality of the integration of both data sets.

Instructional fidelity. Due to the quasi-experimental design of the quantitative portion of my study, I did not have randomly assigned control and treatment groups, instead, I had a nonequivalent control group. Because I wanted to combine the three classes in each semester into a single sample, I had to determine if the three classes displayed instructional fidelity. For my purposes, this means that all the components of the MP lesson were delivered by each instructor. I video-taped all class sessions for each class sections in both the control and treatment semesters. The videos were not used to gather any participant data. I viewed each class in both semesters, and with the aid of coding sheet and recorded when and if the instructor delivered a particular component of the MP lesson (Table 3). The classroom videos demonstrated that for both the control and treatment semesters, the instructors delivered all the components of the MP lesson with

one exception. The MP lesson plan has a hands-on computer activity called BugHunt! This activity demonstrates the selective advantage camouflaged moths have from bird predation. This activity was not included in any class in either the control or treatment semester.

A second aspect of instructional fidelity is concerned with the delivery of the narrative elements by the instructors. The modified scripts contained all ten narrative elements, but this does not mean the instructors followed the scripts. The videos from the Spring semester were viewed with these elements in mind, and the occurrence of each element was recorded. The presence or absence of these elements does not affect my ability to use the three classes as a single sample, the main influence would be on the study's results.

***Validity and reliability*.** For my quantitative data, I was concerned the validity and reliability of my research design and the instrument used. There are three main dimensions of validity that I needed to consider: (a) measurement validity, (b) internal (causal) validity, and (c) generalizability (Shutt, 2009). It is important that the instruments used to measure the study's variables are actually assessing what they are meant to; and therefore I have considered the measurement validity of the CINS. As mentioned above in this chapter, Anderson et al. (2002) reported on the face validity and internal validity of the CINS. Face validity is a type of measurement validity that on inspection verifies the test item does measure the intended concept. The internal validity in this situation is not related to causality, but is concerned with the CINS's underlying measure structure. As reported above both types of validity were verified.

Table 8

Example of Alignment of the Question Numbers, Answer Choices and Alternative Conceptions between the Original CINS Instrument and the Current Version for two Concepts on Variation.

Concept[a]	CINS V3 Definition[a]	Alternative conceptions[a]	CINS 2013 Definition[b]	Alternative conceptions[c]
Variation	Individuals of a population vary extensively in their characteristics. 9d, 16c	All members of a population are nearly identical. 9a, 16a Variations only affect outward appearance, don't influence survival. 9b, 9c, 16b Organisms in a population share no characteristics with others	No two individuals are exactly the same; rather, every population shows enormous variability. 5d, 15d	All members of a population are identical. 5a, 15a All members of a population are nearly identical and variations does not influence survival. 5b, 15b & 5c, 15c
Origin of variation[d]	Random mutations and sexual reproduction produce variations; while many are harmful or of no consequence, a few are beneficial in some environments. 6b, 19c	Mutations are adaptive responses to specific environmental agents. 6c, 15c, 19d Mutations are intentional: an organism tries, needs, or wants to change genetically. 6a, 6d, 19a, 19b	New variation appears randomly through mutation and sexual reproduction.* 6b, 16b	Mutations are adaptive responses to specific environmental agents. 6c, 16c Mutations are intentional: an organism tries, needs or wants to change genetically. 6a, 6d, 16a, 16d

Note [a] Adapted from Anderson, Fisher & Norman (2002)
Note[b] Adapted from Evans & Anderson (2013)
Note[c] Adapted to CINS 2013 from Anderson, Fisher & Norman (2002)
Note Concept included in CINS because it is essential for the process of natural selection, however it occurs before natural selection. (Anderson et al., 2002)

Table 9

Definition of the A Priori Codes Applied to the CINS Instrument and the Interview Questions

Concept	Code	Definition[a]
variation 5,15	SAME[1] 5a, 15a	Individuals are identical.
	SAME[2] 5b,c, 15b,c	Individuals are nearly identical, and/or any variation in traits does not affect survival.
	SCI 5d, 15d	No two individuals are exactly the same; rather, every population shows enormous variability.
variation origin 6,16	ENVCHG 6c, 16c	Specific environmental agents cause direct changes in an individual's traits.
	INTCHG 6a,d, 16a,d	Intentional change: Individuals want, need or try to change genetically.
	SCI 6b, 16b	New variation appears randomly through mutation and sexual reproduction.
variation inherited 7,17	ACQ 7a, 17a, 17d	Changes in an individual's physical traits and behaviors that happen during an individual's life are inherited by offspring.
	BENFIT 7b, d, 17b	Only beneficial traits are inherited by offspring.
	SCI 7c, 17c	Much of the new variation is heritable.
differential survival 8,18	FIT[1] 8a, c, 18a, c	Fitness equated with strength, speed, intelligence, and/or longevity.
	FIT[2] 8d, 18d	Fitness equated with many mates.
	SCI 8b, 18b	Survival in the struggle for existence is <u>not random</u>, but depends in part on the hereditary constitution of the surviving individuals. This unequal survival constitutes a process of natural selection.
change in population 9,19	POPCHG 9d, 19d	Genetic change/mutations occur to meet the needs of the population in a changing environment.
	ALLCHG 9a, 19a	Genetic change occurs gradually in all individuals of the population.
	9c, 19c [b]	Not included
	SCI 9b, 19b	Over the generations the process of natural selection will lead to a continuing gradual change of populations, that is, to evolution. 9b, 19b
	OTHER	Explanation/part of explanation does not correspond with any of the above codes. Explanation is for a different concept.
	NOANSWR	No answer is given

Note [a] Codes are based on accepted explanations and alternative conceptions used for distractors in Anderson et. al (2002)

Note[b] Code for 9c and 19c are problematic as they appear to be about inherited variations. These codes differ from Anderson et. al (2002). This issue is discussed in detail in the body of the paper

Internal validity is concerned with causality, which determines if the variation in one variable results in variation of a second variable. Quasi-experiments are interested in internal validity. There are five criteria that are important in establishing causal relationships between variables. First, there is the association between the independent and dependent variables. It is possible for quasi-experiments to provide unambiguous evidence of association. Second, there are the time order effects of one variable on another. Nonequivalent designs, those with a control and treatment group, may not meet this criterion because it cannot be known for sure that an unknown feature of the group is not what attracted the participant to it. However the use of a pre-test can help determine time order effects of the independent variable on the dependent variable. Third, there is the need for nonspurious relationships between variables. Quasi-experimental studies, while not certain to rule out all extraneous factors, by adhering to proper design protocol, can minimize most external influences. The fourth criterion is identifying a causal mechanism. The research designs of nonequivalent studies do not lend themselves to the identification of a causal mechanism. Fifth, the context of the study can provide a understanding of the causal relationship between the variables. Quasi-experiments that have multiple groups may be able provide information about context. Of the five criteria for causality, nonequivalent studies can meet three of them; association, time-order effects, and nonspurious relationships. Therefore, depending on the results, it may be possible to make claims about causal relationships (Shutt, 2009).

My research design has two specific threats to internal validity. The first is selection bias. Selection bias is a type of internal invalidity that occurs when the control and treatment groups vary in their characteristics. The threat in nonequivalent groups is

very high. This threat can be managed by the use of pre-tests, which my research design does include. Demographic data can also help determine how comparable the groups are. I collected information on gender, age and race from all classes to confirm that the aggregated samples were similar. The results of the demographic data are in Table 9. This table demonstrates that for all three criteria considered there was no significant difference between the control and treatment group. In addition to the demographic data, I compared the participants' aggregated final exam scores and course GPAs and for both groups. The mean for final exam scores in the control group was 74.65, and for the treatment group it was 72.65. The independent t- test had a p-value of .20. The mean for the course GPA in the control group was 84.92, and for the treatment group it was 86.16. The independent t-test had a p-value of .19. Both the final exam scores and GPAs between groups demonstrated no significant difference. With the results of the demographic information, along with the final exam scores and course GPAs, I can be reasonably sure that these two groups are comparable, and do not suffer from selection bias.

Table 10

Demographic Data for the Control and Treatment Groups

Demographic information	Control	Treatment	Pearson Chi-square
Gender			
Female	85%	89%	
Male	15%	11%	P = 0.4
Age			
18-23	90%	87%	
24+	10%	13%	P = 0.5
Race			
Caucasian	88%	87%	
African-American	7%	9%	
Other	5%	4%	P = 0.8

The second threat to internal validity is endogenous change. This type of internal invalidity happens when natural changes involving a participant affect the results. One kind of endogenous change is testing. The taking of a pre-test can make participants aware of information that can cause them to respond differently on the post-test. The use of a control and treatment group mitigates this issue because both groups would be equally affected (Shutt, 2009). As mentioned above, my research design has a control and treatment group and therefore the threat of endogenous change through testing. In summary, my use of pre and post-tests, control and treatment groups, demographic information and academic data keeps the threat to internal validity to a minimum.

The third dimension of validity is generalizability. There are two types of generalizability, sample and cross-population also known as external validity. Sample generalizability means the research results of a sample can be applied to the entire population. External validity means the research results about one group can be applied to other groups. My participants were obtained through convenience sampling. Because I have a nonequivalent research design, my results are not generalizable to the entire student population in the BIOS 1700 courses or to other groups (Shutt, 2009). In summary, my research design can be evaluated for measurement and internal validity, but the generalizability of my study is limited. In addition to validity, the reliability of my research must be considered.

The reliability of an instrument measures the consistency of the participant scores when there is no change in the phenomenon. Reliability is important because it is a requirement for measurement validity. Interitem reliability, also known as internal consistency is important to my research because it is concerned with the consistent

measurement of multiple items that are highly related to each other. The CINS uses 20 items to measure concepts of natural selection. The internal consistency of the CINS was originally measured by Anderson et al. (2002) and found to be sufficient. However, it is important to measure internal consistency with each participant sample. I evaluated the internal consistency of both the control and treatment group by applying a well-known measurement, Crombach's alpha (Shutt, 2009).

***Trustworthiness*.** There are a variety of viewpoints on the concepts of validity in qualitative research, but generally validity refers to the concepts of trustworthiness or credibility, which are how well the researcher conveys the participants' experiences and viewpoints (Lincoln & Guba, 1985; Tashakkori & Teddlie, 2008). There are several strategies for ensuring the trustworthiness of qualitative research. During the research process I employed several strategies to safeguard the quality of my findings. The first strategy is triangulation which means gathering data from multiple methods and/or sources. I did use multiple methods of data collection and multiple sources of data. I also engaged in peer debriefing throughout the research process in my weekly research group meetings. Peer debriefing is the act of sharing findings with critical allies and considering their feedback in the research process. During data analysis, I tried to be aware of the presence of disconfirming evidence, and finally, members of my research group helped me develop and test my codes. This process is known as inter-coder reliability (Marhall & Rossman, 2011). I managed researcher bias by noting how my approach to collection, analysis and interpretation of the data could have been influenced by my attitudes and past experiences (Creswell, 2007).

***Mixed methods*.** The inferences generated from the integration of both data sets should be evaluated for integrative efficacy. This means that the inferences from both the quantitative and qualitative data sets have to be adequately incorporated into the meta-inferences (Tashakkori & Teddlie, 2008).

Limitations. My research design has the following limitations. First, as mentioned above, the use of a convenience sample limits the generalizability of the results. However, the results which can be subjected to statistical testing, can consider causal relationships. Any relationships found can be considered in subsequent studies which could build a case for generalizability. Other limitation of my study is the decision not to test the participant's retention of knowledge at some point in the future. Two factors influenced my decision. First, tracking the participants for a follow-up test would be difficult. Second, time is limited and this additional portion is beyond the scope of this current study.

CHAPTER IV
RESULTS AND DISCUSSION

In the previous chapter I discussed the materials and methods used to conduct my research in order to answer the research questions and to ensure methodological integrity. This chapter reports and discusses the results of my research by considering each research question individually. Each section states the relevant research question and then provides a brief recap of the methods before reporting and discussing the results.

Research Question 1

RQ1. What differences in MPU learning outcomes are identified in participants' CINS scores in both the traditional historical and the story approach?

CINS Results. As discussed in Chapter three, I used the CINS instrument as a pre and post-test in both the control and treatment groups to measure the participant's understanding of natural selection concepts. I evaluated the CINS results using both descriptive and inferential statistics. For the descriptive statistics I found the mean (average) and standard deviation for the pre-and post-tests of each group. With the descriptive statistics I evaluated the CINS results in three additional ways. First, I looked at the average CINS scores for each group, and they were used to determine if (a) there was significant change within and between the groups, and if so, the nature of that change, and (b) if there were significant change in explanatory coherence within and between the groups, and if so, the nature of that change. Second, I looked at the participants' individual CINS score to determine (a) the percentage of participants that demonstrated change within each group, and (b) the percentage of participants who met

specific target scores, and (c) the percentage of participants that demonstrated change in explanatory coherence within each group. Third, I considered the scores for each question and for each pair of concept questions to determine if there was significant change within and between groups.

Average CINS scores. As discussed in Chapter Three, for both the question scores and the concept scores, I calculated the values for all twenty questions and then for ten targeted questions. I applied the descriptive statistics to the average CINS scores and found the control group moved from 51% to 53% on the pre and post- test; and the treatment group moved from 53% to 62% To determine if there were any significant changes in the CINS scores between the pre and post-test for both groups, I performed a paired t-test to the participants' average scores. The results for the control group show that for the 20 questions no significant gains were made with $p = .13$, but for the 10 targeted questions significant gains were made with $p = .04$ (Table 11). For the treatment group, significant gains were made for both the 20 questions and the 10 questions, with $p \leq .001$ for both. I performed an independent t-test to determine if there were significant differences between the control and treatment groups. There was a significant gain between groups for the 20 questions with a p value of .03 but not for the 10 targeted questions between with $p = .07$ (Table 11). There are significant gains between the groups in the 20 question format but not in the 10 question format. It is noteworthy that the story version used in the treatment group had a significant positive impact on both question formats but the traditional version only on the 10 question format.

The two versions of the MP lesson varied only in addition of the narrative elements in the story version, which means the context of the story version is different

from the traditional version. Constructivist learning theory suggests that learning and knowledge retrieval is context dependent and learners actively construct knowledge through social interaction and their experiences (Driver & Oldham, 1986). Therefore it is reasonable to suggest that the context differences in the story version provided the treatment group with alternative interactions and experiences of the MP lesson compared to those in the control group. These alternative experiences may account for the difference in statistical significance between the groups in the 20 question format. The 10 question format did not display the same significance between groups because both groups demonstrated significant change within each group. This means that both versions of the intervention were successful in conveying the five concepts targeted in the MP lesson, but Table 11 shows that the story version had greater gains.

Table 11

Results for Descriptive and Inferential Statistics of the CINS Scores in the Control and Treatment Groups for the 20 and 10 Question Format

CINS	20 Questions				CINS	10 Questions			
	Control		treatment			control		treatment	
	Pre	Post	Pre	Post		Pre	Post	Pre	Post
M	10.20	10.66	10.59	12.30	*M*	4.68	5.29	4.57	5.91
SD	3.19	3.31	3.38	4.54	*SD*	2.06	2.00	1.72	2.72
	control		treatment			control		treatment	
<G>	.46		1.7		*<G>*	.61		1.35	
<g>	.05		.18		*<g>*	.14		.89	
d	.14		.43		*d*	.30		.59	
P .05	.13		< .001		*P .05*	.04		< .001	
	control vs. treatment					control vs. treatment			
P .05	.03				*P .05*	.07			
d	.42				*d*	.32			

While it is important to determine if any significant change in scores have happened, this does not tell us about the nature of the change.

To determine the nature of the change, I applied two different metrics. First, I calculated normalized gains within each group (Table 11). Normalized gains (<g>) is a measurement of instructional effectiveness that compares the maximum possible gains to average gains (Hake, 1998, 2002). For all 20 questions the control group and the treatment group had <g> values below 0.3, which are considered low. For the 10 targeted questions, the control group again demonstrates low gains, while the treatment group has a <g> value above 0.7 which is considered high (Hake, 1998). This suggests that the treatment group's instruction for the MP lesson was more effective than the control group for the concepts taught in that lesson. The treatment group had a <g> of 0.89 which means the participants approached the maximum gains available.

Next, I calculated the effect size within and between groups (Table 11). The effect size is a measurement of the magnitude of difference between groups. For all 20 questions the treatment group demonstrates a greater magnitude of difference then the control group, this is also the case for the 10 targeted questions. Cohen's interpretation of the effect size value lies on a continuum from 0.2 to 0.8, with 0.2 being small, 0.5 moderate and 0.8 being large. However, Cohen cautions that the arbitrary use of this scale may not be suitable for all research (1988). Education research often has small effect sizes (Valentine & Cooper, 2003) and that effect size values of 0.2 have been considered important when based on measures of academic achievement (Hedges & Hedberg, 2007). However, John Hattie's (2008) acclaimed book on over 50,000 studies, *Visible Learning: A Synthesis of 800+ Analyses on Achievement,* states that an effect size of 0.4 is the threshold for desired effects. For example, Hattie (2008) determined that class size has an effect size of 0.2, which according this threshold is not practically

significant and therefore efforts concentrated on changing class sizes should not be undertaken, The threshold of 0.4 chosen by Hattie (2008) has faced criticism. In one review, Snook, O'Neill, Clark, O'Neill and Openshaw (2009) discuss several concerns, based on Hattie's methodology, about choosing a benchmark for effect sizes. For example, Snook et al. (2009) state that an effect size of 0.2 in a study that has a large, randomly chosen sample, which increases reliability and validity, would be considerably more significant than one with a small sample size. They argue that the quality of the studies used in the meta-analysis is important, and one that does not exclude inadequate studies is misleading. This is a concern because Hattie (2008) acknowledges that his study did not discuss the quality of most of the research in the meta-analyses. Another concern was voiced by Arnold (2011) who notes that Hattie did not use weighted averages to evaluate the 800 meta-analyses, which assumes that meta-analyses containing a large number of studies hold equal value to those with small numbers of studies. As with Cohen's scale for effect size, the cautions expressed above suggest that Hattie's threshold of 0.4 should not be applied to all educational studies.

It is desirable to have effect sizes from similar studies to compare values against (Valentine & Cooper, 2003). Comparing effect sizes among similar studies provides a normative comparison of those studies, which means that a study that demonstrates the largest effect size within a similar group of studies has practical significance by virtue of that effect size, regardless of its actual value. (Lipsey, Puzio, Yun, Hebert, Steinka-Fry, Cole, Roberts, Antony & Busick, 2012). I feel that comparing this study's effect size to the effect sizes of similar studies is more useful than using a threshold value as discussed above. I found three studies on teaching evolution and/or natural selection that reported

effect size using Cohen's d. One study supported by Biological Science Curriculum Study (BSCS) on using a medical case study to teach evolution had a Cohen's d of 0.20 on the natural selection portion (Beardsley, Stuhlsatz, Kruse, Bllom & Westbrook, 2011). Another study which using a computer -based laboratory intervention to alleviate student misconceptions about natural selection demonstrated a significant drop in misconceptions from the pre to post-test, but Cohen's d was 0.059 (Abraham, Meir, Perry, Herron, Maruca & Stal, 2009). A study on the acceptance of evolution and the change in misconceptions after a 2-hour lesson, has a Cohen's d of 0.51 and 0.19 in the two trials (Abraham, Perez, Downey, Herron & Meir, 2012). While all these studies had mainly small effect sizes according to Cohen's traditional scale, all studies had statistically significant change in the pre to post-tests. The effect sizes for the treatment group in my study are comparable or larger than these three studies. Effect size is one to determine if the results of a study have a practical significance (Maher, Markey, & Ebery-May, 2013), and since educational research tends to generate small effect sizes the benefits of these studies should not be under-valued.

When all three metrics applied to the participants' average scores are considered within each group, we can see that the treatment groups' values are consistent with larger gains from the pre to post-test compared to control group. The treatment group's results demonstrate: (a) significant gains in scores for both the 20 and 10 question format, and (b) higher normalized gains and effect size values. The results also show that the treatment group scores were statistically significant higher than the control group in the 20 question format. In spite of the treatment group's gains, the average CINS post-test scores for both question formats were ~ 60%. While these averages are not as high as

hoped for; they are not unexpected. Many previous studies have reported encouraging results after the intervention are still below desired targets (Bishop & Anderson, 1990; Jenson and Finley, 1996; Nehm and Reily, 2007; and others).

I next used the average CINS scores to investigate explanatory coherence amongst the participants. As mentioned earlier in Chapter 3, explanatory coherence is how consistently participants answer different questions on the same concept, which is an indication of whether the instruction was effective or not (Kampourakis, 2009). To determine explanatory coherence, I paired the CINS questions into the 10 concepts on the instrument. I determined if the participant responded correctly to both questions for each concept, if not they did received a zero for that concept. Therefore, participant scores for explanatory coherence could range from zero to ten. I also determined explanatory coherence for the 5 concepts pertaining to the MP lesson, and those scores ranged from zero to five. The scores for participants were averaged and the results for explanatory coherence are in Table 12.

The results for the t-tests demonstrate: (a) that within the control group for both the 10 and 5 concept pairs, no significant gain was made from pre to post-test, (b) in the treatment group there were significant gains made from pre to post-test in both the ten and five concept pairs ($p < .001$ for both), and (c) there were significant gains made between the control and treatment group in both the ten and five concept pair ($p = .03$; $p = .007$). This means that the gains in explanatory coherence in the treatment group and between the control and treatment group did not happen by chance. Unlike the average scores discussed above, the story version used by the treatment group displayed significant positive gains in all measurements and the control group did not.

Table 12

Explanatory Coherence for the 10 and 5 CINS Concept Pairs in the Control and Treatment Group

CINS Scores	10 concept pairs				CINS Scores	5 concept pairs			
	control		treatment			control		treatment	
	pre	post	pre	post		pre	post	pre	post
M	3.44	3.63	3.76	4.74	*M*	1.61	1.90	1.54	2.30
SD	2.10	2.05	2.07	2.65	*SD*	1.26	1.22	1.03	1.59
<G>	.20		.98		*<G>*	.30		.76	
<g>	.03		.16		*<g>*	.09		.22	
d	.09		.41		*d*	.30		.60	
P .05	.24		< .001		*P .05*	.09		< .001	
	control vs. treatment					control vs. treatment			
P .05	.03				*P .05*	.007			
d	.42				*d*	.35			

To understand the nature of this change, I calculated both normalized gains and effect size. The normalized gains for the control and treatment group for both the 10 and 5 concept pairs are considered low with all values < .30 (Hake, 1998). This means that while the participants in both groups made gains between the pre and post-tests, the gains achieved were not near the maximum available. The results for effect size demonstrate that the Cohen's d was higher for the treatment group in both the ten and five concept pairs. The treatment group versus the control group demonstrates effect sizes comparable or higher to other research on evolution and/or natural selection. While there were significant gains in the treatment group, the scores remained < 50%. This means on average, the participants only achieved explanatory coherence on half of the concepts. The science education literature on explanatory coherence is limited, however in one study, Kampourakis and Zogza evaluated middle school students on five tasks that covered the concepts of common descent and natural selection. They determined a

student had achieved explanatory coherence if they gave appropriate explanations for all five tasks. About one third of the students achieved coherence after instruction (2009). This method of evaluation is not applicable to my study because the participants were evaluated on 10 concepts covered by 20 questions. No one earned 100% on the CINS and therefore according to the Kampourakis and Zogza method no one achieved coherence. Instead, I measured how many of the 10 or 5 target concepts the participants achieved coherence in and took the average. In this way gains could be measured. To approximate the Kampourakis and Zogza method, I determined which concepts had the highest percentage of participants achieve coherence; this is reported later.

Individual CINS scores. I evaluated the participants' individual scores to determine: (a) the percentage of participants in each group that demonstrated a positive, negative, or lack of change, (b) the change in the percentage of participants from the pre to post-test in each group that reached specific grade benchmarks (Table 13), and (c) the percentage of participants in each group that demonstrated a change in explanatory coherence (Table 14). First, the results for changes between the pre and post-test demonstrate that for the 20 and 10 question formats, the treatment group had a higher percentage of participants that gained from pre to post-test. The values for both formats are very similar, with the control group about 50% and the treatment group about 60%, however only the 10 question format showed significant gains from control to treatment in the z-ratio analysis ($p = .04$). The losses experienced by the control group and treatment group in both formats were about 30%; with the exception of the 10 question format for the treatment group, this was about 20%. The percentage of participants in the control and treatment groups with no change were similar with the exception of the 20

question format for the treatment group which was only 4%. The difference between the control and treatment group for "no change" in the 20 question format was significant ($p < .001$).

My next task was to determine the change in the percentage of participants that earned at least a 60% and those that earned at least a 70%. I chose the 60% benchmark because it is usually the minimum for passing, and the 70%, because it is usually the percentage needed for transferring a grade. Table 14 displays both the percentage of change in the control and treatment groups for the 20 and 10 question formats.

The 20 question format shows that the control group had 71 % of participants earn < 60% in both pre and post-test. The only change was 5 % of participants that moved from the 60-69% group to the ≥70% group. In contrast, the treatment group had a significant change of 26% of participants move out of the < 60% group. The majority (22%) of these participants moved to the ≥ 70% group. The control group demonstrated

Table 13

Percentage of Participants with CINS Score Changes from the Pre to Post-test

CINS Scores % of participants	Q20		Q10	
	control	treatment	control	treatment
gain	54	63	51	63
loss	32	33	29	22
no change	15	4	20	15

some improvement in the 10 question format. 10% of participants moved out of the < 60% with most (7%) going to the 60 - 69% group. The treatment% group again

demonstrated change (33%) out of the < 60% group to the ≥ 70% group. In addition, there was a 2% change from the 60 - 69% group to the ≥ 70% group.

In the control group, for both formats, the following results were observed. First, the percentage of participants in the < 60% group remained over 50% Second, less the 25% of participants earned scores ≥ 70%. In the treatment group, for both formats, ~ 40%

Table 14

Percentage of Participants with Specific CINS Scores on the Pre and Post-test

percentage of participants	Q20					
	control			treatment		
	pre	post	% change	pre	post	% change
< 60%	70.7	70.7	0	67.4	41.3	- 26.1
= 60 - 69%	14.6	9.8	- 4.8	13.0	17.4	+ 4.4
≥ 70%	14.6	19.5	+4.9	19.6	41.3	+21.7
percentage of participants	**Q10**					
	control			treatment		
	pre	post	% change	pre	post	% change
< 60%	68.3	58.5	- 9.8	71.7	39.1	-32.6
= 60 - 69%	9.8	17.1	+7.3	15.2	13.0	-2.2
≥ 70%	22.0	24.4	+2.4	13	47.8	+34.8

of participants remained in the < 60% group and over 40% earned scores ≥ 70%. The classification of participants in grade percentage groups demonstrates a positive trend in the treatment group. It shows that the majority of participants earned failing grades in the pre-test but in the post-test this had shifted to the majority earning over 60%. These results show why it is important to consider the CINS results from the perspective of the individual participant. These results move beyond the information gleaned from the average scores and reveal significant positive participant movement from failing to passing scores.

Next, I looked at explanatory coherence scores for individual participants to determine the type of changes in that happened from pre to post-test (Table 15). The treatment group had higher gains compared to the control group for both the 10 and 5 concept pairs. However, only the 5 concept pairs demonstrated significant difference in the z-ratio analysis with a p-value of 0.02. The treatment group demonstrated lower percentages of participants that experienced losses in their scores. The differences in the losses for the 20 question format between the control and treatment group was significant with a p-value of 0.03.

Table 15

Percentage of Participants with CINS Score Changes for Explanatory Coherence from the Pre to Post-tests

CINS Scores	10 concept pairs		5 concept pairs	
% of participants	control	treatment	control	treatment
gain	44	48	39	54
loss	34	22	24	17
none	22	30	37	28

Question performance. After looking at the participants' scores together and as individuals, I evaluated the CINS scores by question to determine how participants as a group performed on each one. The questions were evaluated in two ways: (a) in the 20 question format, and (b) in the 10 concepts pairs for explanatory coherence. I performed a McNemar's test on both groups' pre and post-tests scores for each question or concept. For each question or concept, I performed a chi-square analysis to determine significance between groups.

For the 20 question format, the treatment group had gains on more questions than the control group, but McNemar's demonstrated both the control and treatment groups had similar numbers of questions that had significant gains from the pre to post-test (Table 16). The concepts involved with these questions are limited resources (#13) and change in population (#9, #19). Change in population is a concept included in the MP lesson. While there were no significant losses for any question in either group, the treatment group had fewer losses than the control group. There were two questions in the

Table 16

McNemar's Results for each of the 20 CINS Questions

CINS Scores # of questions	Q20			
	control		treatment	
	total	Sig.[a]	total	Sig.[a]
gain	9	2	16	3
loss	6		4	
none	5		0	
Significant questions		**#9 #19**		#13 **#9 #19**

control group had two questions with significant gains and the treatment group had three. While there is no significant difference in the number of questions with gains between the control and treatment group, there is an interesting pattern between the groups. The control group had a total of nine questions display positive gains and the treatment group displayed 16. This supports the idea that the story version of the MP lesson used in the treatment group had a global impact on the participants' understanding of natural selection. Finally, the Chi-square analysis for the 20 CINS questions did not demonstrate any significant change between the control and treatment group.

The results for McNemar's test on the 10 concept pairs are displayed in Table 17. This results show that the treatment group has gains on more concepts than the control group, nine compared to four. However, of the nine in the treatment group, only three showed a significant difference and one in the control group, was significant. The concepts that showed significant change were: (a) *population change* in the control and treatment group, and (b) *origin of variation* and *differential survival* in the treatment group. All three concepts are part of the MP lesson. One concept showed a significant loss in the control group, this was *limited resources*. The treatment group showed no losses. In addition to considering what concepts demonstrated change, I looked at what concepts had the greatest percentage of participants achieve explanatory coherence. The control group achieved ≥ 50% coherence on one concept, while the treatment group achieved ≥ 50% coherence on five concepts. The control group had > 50% of the

Table 17

McNemar's Results for Explanatory Coherence in the 10 CINS Concept Pairs

	control			treatment		
type of change	gain	loss	none	gain	loss	none
# of concepts	4 1 sig.	3 1 sig.	3	9 3 sig.	1	0
Significant concept	population change **9,19**[a]	Limited resources 3,13		variation origin **6,16** differential survival **8,18** population change **9,19**		

Note[a] The bolded pairs indict that the significant concept is one of the 5 concepts related to the MP lesson

participants achieve coherence on *variation inherited*. The treatment group had > 50% of the participants achieve coherence of *biotic potential*, *variation* and *change in*

population, and > 60% on *stable populations* and *variation inherited*. Both groups shared variation inherited as a concept that had high participant understanding.

I performed a Chi-square analysis for the 10 concept pairs for the control and treatment groups. The analysis demonstrated that the concepts of biotic potential (1, 11) and stable populations (2, 12) had significant differences. This indicates that the difference in the average participant scores for these two concepts did not happen by chance.

The final aspect of explanatory coherence I wished to examine was the concepts the participants had the most difficulty with. I determined which concepts in each group had less than a 50% correct response rate. The most difficult concepts for the control and treatment groups were the same, with *origin of variation* and *change in population being the most challenging,* followed by *differential survival* (Figures 3 & 4). Generally, less than 50% of the participants in both groups scored correctly on the post-test for the

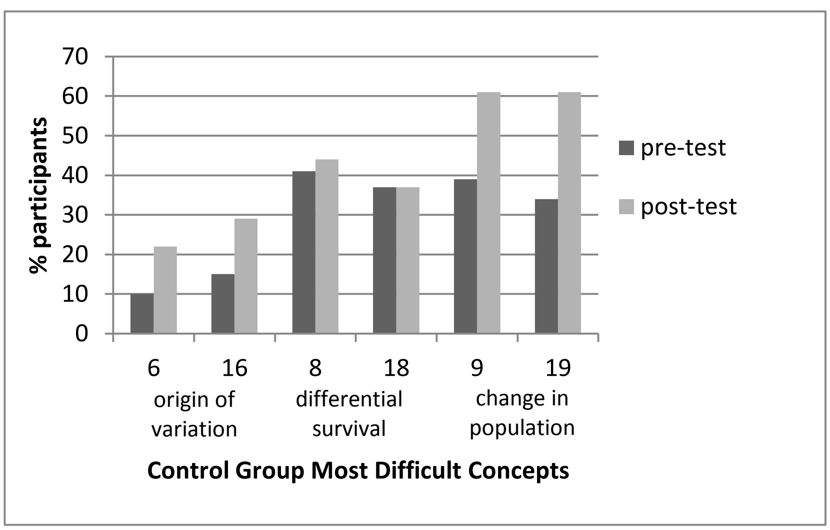


Figure 3. The control group's most difficult concepts are determined by < 50 % of participants receiving a correct answer on both questions in the concept pair.

concepts of *origin of variation* and *differential survival*. The participants in both groups improved the most on the *change in population* concept, with ≥ 60% of each group receiving correct answers. The literature supports that the concepts of *origin of variation*, *change in population* and *differential survival* are problematic to students.

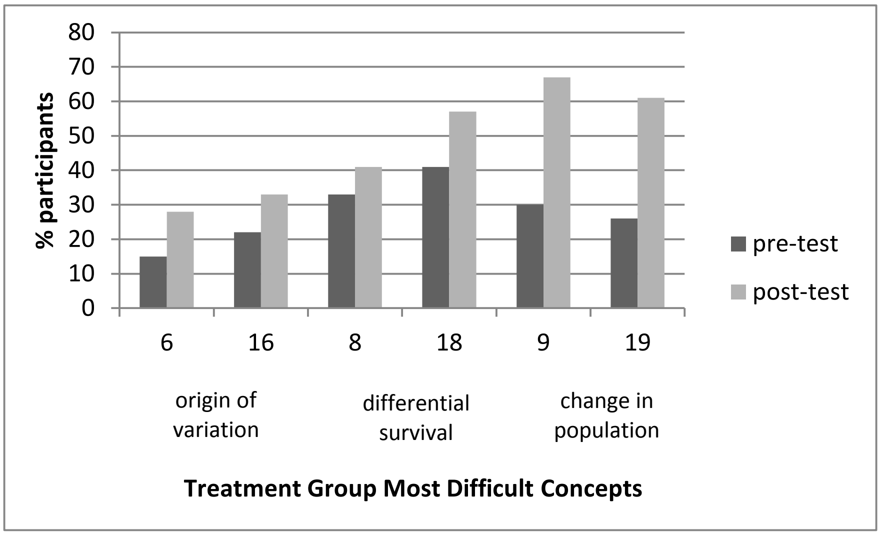


Figure 4. The treatment group's most difficult concepts are determined by < 50 % of participants receiving a correct answer on both questions in the concept pair.

When searching the literature on which concepts were the most difficult for other research using the CINS instrument, I focused on the five targeted concepts in my study. While Anderson was developing the original CINS instrument, she determined the percent of correct responses was well under 40% for *change in population* and *origin of variation,* and *differential survival* and *variation inherited* just under 50% (Anderson et. al, 2002). The results were the same a decade later when the revised version of the CINS was being tested on three different groups of college students (Evans & Anderson, 2013). Nehm and Schonfeld found that found that the concept of *change in population* was the most difficult, followed by *origin of variation* with two groups of biology majors at an urban university in the northeast US (2008). These results were also reported by Battisti

and his colleagues, who also used the original CINS with over one thousand participants in an introductory college biology course (Battisti, et al., 2010). Studies that did not use the CINS instrument also reported difficulties with the concepts of *origin of variation* and *change in population* (Bishop & Anderson, 1990; Jensen & Finley, 1996). So it appears, that the three most difficult natural selection concepts of this study's five targeted concepts, are found to be widely problematic and impervious to instruction.

Nehm and Schonfeld found that found that the concept of *change in population* was the most difficult, followed by *origin of variation* with two groups of biology majors at an urban university in the northeast US (2008). These results were also reported by Battisti and his colleagues, who also used the original CINS with over one thousand participants in an introductory college biology course (Battisti, et al., 2010). Studies that did not use the CINS instrument also reported difficulties with the concepts of *origin of variation* and *change in population* (Bishop & Anderson, 1990; Jensen & Finley, 1996). So it appears, that the three most difficult natural selection concepts of this study's five targeted concepts, are found to be widely problematic and impervious to instruction.

S**upplementary material.** In addition to the participants CINS scores, I evaluated Group G questions on the course's final to determine if there was any significant differences between the control and treatment group. The seven Group G questions were explicitly about industrial melanism and the peppered moth. According to the independent t-test for these two samples there is a significant difference between the scores; with the p-value being 0.02. The effect size is 0.27 which is considered small by the traditional scale given by Hake, but see previous discussion above.

In summary, CINS scores demonstrated that the treatment group (a) had significant gains in average scores, and (b) significant gains in explanatory coherence in both the 20 and 10 question format. The control group had significant gains in average scores for the 10 question format. In addition, the Group G questions on the final exam demonstrated significant gains for the treatment group. In spite of the significant gains in the treatment group, the average CINS scores also demonstrated that after instruction participants in both groups did not provide scientific explanations for most of the CINS questions, or achieve explanatory coherence for most concepts. However, when the individual CINS scores were analyzed, they exhibited a positive occurrence. A significant percentage of the treatment group moved out of the 60% group to the ≥ 70% group in the 20 and 10 question formats. The control group had a significant change out of the 60% group. This means a large portion of these participants moved from a failing grade to not only a passing grade, but to one that is transferable. So while the total percentage of participants in the ≥ 70% group remained under 50%, this was double the percentage in the control group (24%). The control group had a significant change out of the 60% group in the 10 question format. The treatment group out-performed the control group in average scores, individual scores and by question performance in both question formats. The 20 question format results demonstrate that the story version of the MP lesson (treatment group) had a greater impact on participants' overall understanding of natural selection then the traditional version (control group). The 10 question format results demonstrate that both groups had learning gains in the concepts specifically targeted in the MP lesson, however, the story approach (treatment group) gains were more significant.

Research Question 2

RQ2. What alternative explanations, as identified in the CINS distractors, and the interviews are participants using in both the traditional historical and story approaches?

As demonstrated in RQ1, participants were generally not providing scientific explanations for most questions on the CINS instrument. Therefore it is of interest to see what explanations they did choose. As discussed in Chapter 3, the CINS instrument uses alternative conceptions as the distractors in the forced choice questions. I wanted to determine what alternative conceptions the participants were choosing the most often and

Table 18

Participants' Most Frequent Distractor Choice for each Concept

Concept	Question pair	Control		Treatment		A Priori Codes[a]
		pre	post	pre	post	
variation	5	b	b	b	b	SAME[2]
	15	c	c	c	c	SAME[2]
origin of	6	a	a	a	a	INTCHG
variation	16	a	a	a	a	INTCHG
variation	7	d	d	b	d	BENFIT
inherited	17	a	a	b	a	ACQ
Differential	8	a	c	c	a	FIT[1]
survival	18	d	d	d	d	FIT[2]
Change in	9	d	d	d	d	POPCHG
population	19	d	d	d	d	POPCHG

how consistent the choice was within and between the groups. In Chapter Three I discussed the a priori codes that were developed from the alternative conceptions. cons The distractors and their assigned codes are in Table 9. Recall that each concept on the CINS instrument has two questions that pertain to it, so as displayed in Table 18, questions 5 and 15 pertain to the concept of variation. Each concept has two alternative conceptions as their distractors (Table 9).

Alternative conceptions. Results for question RQ2 (Table 18) demonstrate that the control and treatment groups show consistent choice of alternative explanations for all questions in the pre and post-tests. For the concept of *variation*, both groups chose b for question 5 which was assigned the $SAME^2$ code. This code refers to the alternative conception that individuals are nearly identical and differences do not affect survival. This represents a misunderstanding of the role of alleles on variation. For the concept of *origin of variation* the most common distractor choice (a) represents the Lamarckian idea that organisms change because of a need or want. The distractor choices for the concept of *variation inherited* also represent Lamarckian ideas. The code BENFIT refers to only traits that benefit the individual being passed down, and ACQ, represents the idea that traits acquired during an individual's life can be inherited by the offspring. The distractor choices for *differential survival* (FIT^1 and FIT^2), represent the idea that physically or intellectually superior individuals, or those that have more mates, respectively, are more fit. These are misunderstandings about Darwin's concept of fitness, which is concerned with the reproductive success of organisms and the subsequent passing down of specific alleles to the offspring. The distractor choices for the concept of *change in population* refers to the genetic change happens to meet the needs of the population in a changing environment.

While there was remarkable consistency within and between the groups for each question; when each concept was considered as a whole the use of alternative conceptions was not as constant. In both groups, for the concepts of *variation inherited* and *differential survival*, the first question for each concept has one code, and another code for the second question. For the concept of *variation inherited*, the code chosen for

question 7 was BENFIT; only beneficial traits are passed on. The code for the question 17 was ACQ; traits acquired in ones' lifetime can be passed on. For the concept *differential survival*, FIT^1 was chosen for question 8 and FIT^2 for question 18. FIT^1 equates fitness with characteristics such as speed, strength, or health/longevity. FIT^2 equates fitness with the ability to have many mates.

In summary, the most common choice of alternative conceptions from pre to post test was 100% consistent, as was the choice from control to treatment group. The choice of two different alternative conceptions for the concepts of *variation* inherited and *differential survival* are the only inconsistencies in explanations within and between the groups. One could argue that the consistency in alternative conception choice could be related to the fact that the distractors for each question only represent two types of alternative conceptions (Table 9). However, the limited choice in distractors does not explain why one type of alternative conception was always the most frequently chosen. This suggests that the favored distractors provide explanations that have wide appeal. The literature is limited on information regarding distractor choice on the CINS instrument, however, Anderson (2003) provides distractor choice information on four of the five targeted concepts and Evans and Anderson on two of the five (2013). Anderson provides data from five classes in three different institutions, one city college and two universities. The most commonly chosen distractor was given for the concepts of *variation*, *origin of variation*, *variation inherited* and *change in population*. For all five concepts in Anderson (2003), the students chose distractors that represent the same alternative conceptions as this study (Table 18). In Evans and Anderson (2013), the data was generated from eight test populations ranging from middle school students to biology majors in college. The

most commonly chosen distractors for the concepts of *origin of variation* and *differential survival* also represent the same alternative conceptions as this study. The consistency of alternative conception choice within this study, and across studies that span over a decade emphasize how resilient these conceptions are. Instruction should explicitly include these alternative conceptions as a part of a well-designed lesson that keeps constructivist learning theories in mind.

MP Lesson misconceptions. As discussed in Chapter Three, both the original and modified version of the MP lesson explicitly considers some misconceptions (alternative conceptions) in the form of past theorists and scientists' ideas about change in organisms and populations. Table 19 compares the misconceptions explicitly addressed in the MP lesson to the corresponding CINS questions and distractors they are found in. Table 19 displays four misconceptions, three attributed to Lamarck and one to Devries. There are three natural selection concepts covered by these misconceptions, *origin of variation*, *variation inherited* and *change in population.* Five CINS questions and 11 distractors correspond to these misconceptions. Since these misconceptions were explicitly discussed in the MP lesson, it is valuable to determine if the participants' use of those misconceptions decreased from the pre to post-test and from control to treatment group. To determine how the participants faired from pre to post-test, I preformed McNemar's test for correlated proportions.

For the Larmarckian concept of *origin of variation*, both groups had a significant decrease ($p < 0.001$) from pre to post-test. The control group decrease went from 68.0% to 55.0% and the treatment group went from 64.0% to 49.0%. For the Larmarckian

Table 19

Misconceptions Explicitly Considered in MP Lesson

Theorist/Scientist	Concept	Misconception	CINS Question & Distractor & Code
Theorist: Lamarck Scientist: Cooke	Origin of variation	Individuals can adapt to environment by acquiring new traits.	Mutations are intentional: an organism tries, needs or wants to change. 6a, 6d, 16a, 16d INTCHG
	Variation inherited	Acquired traits are passed down to offspring.	Traits acquired during an organism's lifetime will be inherited by offspring. 7a, 17a, 17d ACQ
	Change in population	Populations change over time because the individual organisms composing them are adapting to their environment.	Changes in a population occur through a gradual change in all members of a population 9a, 19a ALLCHG
Theorist: DeVries Scientist: Heslop-Harrison	Origin of variation	Variation happens when genes are altered by mutations caused directly by environmental factors.	Mutations are adaptive responses to specific environmental agents 6c, 16c ENVCHG

concept of *change in population* only the treatment group had a significant decrease ($p < 0.001$) of 22.0% to 6.5%. In addition, the z-ratio value displayed a significant change between the treatment group and control group ($p = 0.009$). The other two misconceptions, the Larmarckian idea about *variation inherited* and the DeVries idea of *origin of variation* did not display significant changes in either group. There was only a small percentage of participants that chose these misconceptions. About 10% of participants in both groups chose the Lamarckian concept of *variation inherited* and only 5% of participants in both groups chose DeVries' *origin of variation* on the pre-test. The

post-test responses for the Lamarckian concept of *variation inherited* increased slightly (control (1%) and treatment (3%)), and for DeVries' *origin of variation* the percentage did not change for the control group, and decreased 1% for the treatment group. The lack of significant change may be due to the small percentage of the participants who chose these misconceptions held them in a tenacious fashion.

The MP lesson included four misconceptions, the CINS results show that one of those, the Lamarckian concept of *origin of variation* was chosen by over 60% in both groups in the pre-test and this dropped to about 50% on the post-test. The significant change in both groups is clearly a case of the MP lesson explicitly targeting a common misconception and positively affecting its decline. Of the other three misconceptions explicitly included in the MP lesson, the Lamarckian concept of change in population was chosen by ~ 20% of the participants and only the treatment group displayed significant change and there was a significant difference between the control and the treatment group. The control group did not display significant change for this concept, however the treatment group did. This lends support to the idea that the story version of the MP lesson has a positive impact on this misconception. The last two misconceptions did not display significant change or have a lot of participants choose them, so with this group of participants, the explicit inclusion of these misconceptions did not have significant impact on the CINS outcomes.

This portion of the research question revealed that the participants in this study chose the same alternative conceptions as those in other studies. In addition, the MP lesson explicitly targeted four misconceptions, two of which represented the most commonly chosen distractors in three CINS questions (6, 16, and 17). These

misconceptions were the Lamarckian concepts of *origin of variation* (INTCHG) and *variation inherited* (ACQ) see Table 18. In the original version of the MP lesson the use of one misconception displayed significant improvement and two displayed significant improvement in the modified (story) version. These improvements are encouraging, and in future versions of the MP lesson more of the common misconceptions used by students should be explicitly targeted.

As mentioned above, the CINS revealed that the participants in this study not only held the same alternative conceptions as each other, they held the same conceptions as participants in other studies. This makes the CINS a useful instrument for evaluating the efficacy of interventions. However one limitation of the CINS instrument is that there are only specific alternative conceptions available for the participants to choose. Participants could hold alternative conceptions outside the range of those found on the CINS. In addition, the participants' understanding of the concepts of variation and natural selection may be more complex that revealed in the CINS responses. Therefore, a subset of participants were interviewed about their understanding on the concepts of variation and natural selection.

While the CINS instrument demonstrated how the participants scored on specific questions, the open-ended nature of the interview questions gave the participants the opportunity to discuss any aspect of both concepts. The interview questions had two foci. The first was the main purpose of the lesson; IQ1 and IQ4 asked the participants to identify and talk about the main point of the lesson and about the mystery that was being explored in the lesson. The second focus was on the participants' specific explanations about the concepts of variation and natural selection. Questions IQ2 and IQ3 asked:

“what is variation”, “How does variation happen?”, “What is natural selection?”, and “How does natural selection work?” Because the questions are open-ended, the participants had to decide what the question meant to them and what to share about it. This gave them the opportunity to share what they knew and thought was important. Together the interview questions revealed what the participants thought the MP lesson’s purpose was; and what they knew about the scientific concepts discussed in the lesson.

Main point. The reason for asking interviewees about the main point of the MP was to see if their idea of the lesson’s purpose was aligned with any of the objectives in the lesson plan. The MP lesson plan has multiple objectives for each day, and there was no expectation that the interviewees would touch on all of the objectives, just those that they perceived as the main point. Both groups’ responses for the interview questions on the lesson’s main point (IQ1a, b) and the mystery to solve (IQ4a-c) were very similar and

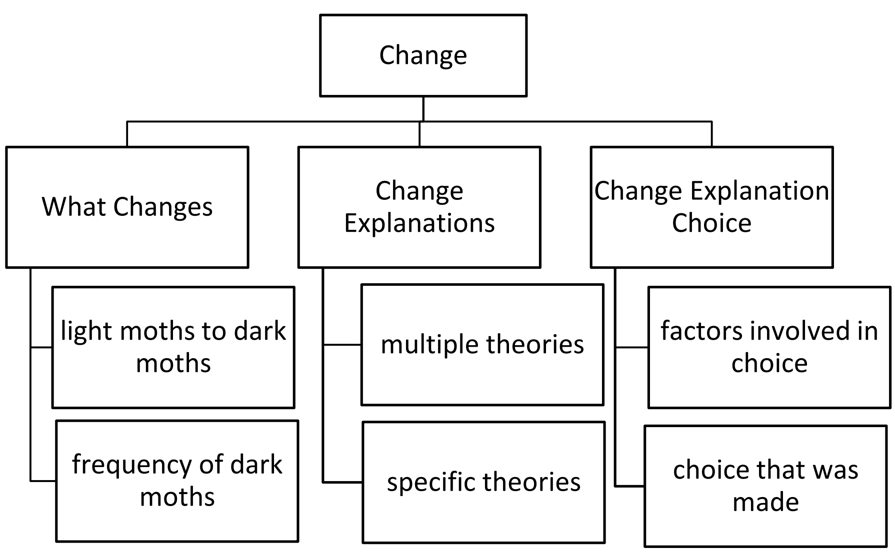


Figure 5: Interview questions 1 and 4 reveals the meta-theme of change.

Therefore, I considered both sets of questions together to form an overall picture of both groups' ideas. One meta-theme, change, emerged from the responses. Figure 5 represents the meta-theme change and its corresponding themes, which are: (a) what changes, (b) explanations of change, and (c) change explanation choice.

What Changes. For the theme of "what changes", most of the responses in both groups involved the peppered moth (Table 20). Interviewees said the lesson's main point was to explain why the light moths became dark, and/ or why the dark moths were increasing in number. About one third of interviewees in both groups stated the purpose was to figure out why the light moths were turning dark. This is an alternative conception about the ability of an individual to change their traits which is related to the concept of *origin of variation*. This is discussed later in the section on variation. Another third of each group said the purpose was to explain why the frequencies of dark moths/moths

Table 20

The What Changes Theme Mainly Involves the Peppered Moths

What Changes	Control n=15	Treatment n=14
Moths		
• Light moth → dark moth	5	4
• ↑ dark moths	5	3
• light→ dark & ↑ dark moths	2	4
• change moth frequency	0	1
General		
• traits, species, animals	2	2
No Answer	1	0

were increasing. A few interviewees included why the moth frequency changed in a specific area, such as industrialized or polluted area and this is the stated mystery at the beginning of the lesson. Some interviewees thought the purpose was to explain why light

moths were turning dark and why the dark moth frequency was increasing. The remainder of the responses said the purpose was to explain how traits, species or animals were changing. The next theme was about explanations for the change in the moths.

Change explanations. For the theme of "change explanations" interviewees talked about the theorists, theories or scientists discussed in the lesson. They said that the lesson was about learning the theories that would explain the changes in the moths. The majority of interviewees in both groups talked about theories in general, and some went on to list or mention about specific theories of change (Table 21). One difference in the groups was that specific theories were listed more often in the control group. A third of

Table 21

The Change Explanations Theme Involves Different Theories of Change

Change Explanations	Control n=15	Treatment n=14
Multiple theories	12	12
Specific theories		
• natural selection	14	5
• mutation theory	9	2
• acquired inheritance	4	1

the interviewees in each group also mentioned evolution in connection to change. They did not talk about evolution as a theory of change, but more as a synonym for change. Addition to talking about the explanation of change, about 2/3rds of the control group gave reasons for the changes in the moths. All these reasons involved survival. Only two interviewees in the treatment group talked about survival as a reason for change. About one half of the control group and one third of the treatment group mentioned that the environment had an influence on the moths changing; however, the responses did not

include specific mechanisms of change. One surprise was the lack of mention of the change in the moths having a genetic component. Only one interviewee in the control and two in the treatment group connected changes to genetics. In general, interviewees remained focused on the names of the theories they had to learn, and did not offer a lot of details about those theories. The third theme that emerged was how the interviewees chose their preferred explanation(s) of change.

Change explanation choice. As mentioned above, the interviewees were asked about the main point of the lesson, and the mystery they were trying to solve. The groups responded that learning about change in the moths and about the explanations responsible for that change were the main points of the lesson. Another thing that both groups talked about were factors that influenced their choice of change explanations. These factors, along with the chosen change explanations are in Table 22.

The majority of both groups personally felt that the MP had been solved during the course of the lesson. About half of the interviewees talked about the fact that there was no definite answer to the MP. Because of this, they went on to explain that picked the answer that made the most sense and/or the best answer. About a third of both groups mentioned that it was the evidence presented in the lesson that helped them pick an explanation. Two interviewees in the treatment group stated that they were confused about the lesson and went with the majority. One interviewee in the control group said the MP was not solved because theories are not facts and therefore no answer could be chosen. In the treatment group, two interviewees said the MP could net be solved because there is no definite answer. Overall, the interviewees who thought the MP had a solution

Table 22

The Change Choice Theme includes the Interviewees' Choice and Factors Involved in that Choice

Change Choice	MP solved		MP Not Solved	
	Control	Treatment	Control	Treatment
Choice personal	12	10	1	3
Factors				
• no definite answer	5	6	-	2
• best answer	2	5	-	-
• makes sense	7	4	-	-
• evidence	4	3	1	-
• majority	-	2	-	-
Theories				
• natural selection	10	7	-	-
• acquired inheritance	1	-	-	-
• mutation	-	-	-	-
• directed change	1	2	-	-
• not sure	-	1	-	-
Choice class	9	10	3	2
Factors				
• no definite answer	1	1	-	-
• best answer	1	1	-	-
• makes sense	1	1	-	-
• evidence	1	1	1	-
		-	-	-
Theories				
• natural selection	7	8	-	-
• acquired inheritance	-	-	1	-
• mutation	-	-	1	-
• directed change	1	-	-	-
• not sure	1	2	-	-

gave reasons for their choices; while those who thought it could not be did not provide reasons for their choice. Of the three theories discussed in the lesson, natural selection was selected as the solution to the MP by most of the interviewees. Only one interviewee in the control group chose Lamarck's theory of acquired inheritance, which was also

presented in the lesson. In addition to the theories discussed in the lesson, a few interviewees mentioned that organisms needed to change in order to survive in a changing environment. This is noted as directed changed in Table 22.

While most interviewees chose natural selection as their preferred change explanation, this choice was not necessarily straight forward. One interviewee, F2 from the control group, felt that the MP did explain how the moths survived in the changing environment, and the theory that did the best was natural selection. However, F2 did not feel that natural selection explained how the change happened in the moths, and they were leaning toward the mutation theory as an explanation. Earlier F2 stated that main point of the mystery was to explain how the light moths turned dark. This misinterpretation of the mystery may be the reason F2 felt they reached only a partial answer. In the treatment group, two interviewees, S10 and S13 felt that natural selection was occurring, but that the presence of the 3rd variation suggested that another process was also contributing to the phenomenon. Neither one was able to express what that process might be. , but they felt there was more to the story than natural selection. Another interviewee, S6, also found the introduction of the 3rd variation problematic. S6 felt that the mystery was solved before the addition of the 3rd variation; and therefore it was not necessary to the lesson. While S6 felt the mystery was solved, they were not able to recall the solution.

The interviewees were asked if they thought the class came to consensus on .the solution to the MP (Table 22). Most of the interviewees in both groups felt their class agreed on a solution. Natural selection was chosen the most often as the preferred solution. Only one interviewee (F3) felt the class chose another explanation, which was

directed change. One interviewee in the control and two in the treatment group felt their class came to consensus, but were not able to say what the preferred solution was. A few interviewees in both groups did not feel their classes came to consensus. In the control group, F7 felt that while the class did not agree on a solution, they did think the Acquired Inheritance Theory and the Mutation Theory had good points. Also, F9 said there was no consensus because they were just talking about theories and not facts, and finally, F12 felt their class was confused and could not come to agreement. In the treatment group, the two interviewees who felt their class did not reach agreement gave no insight on the decision making process. While Interviewees talked about how they made their personal decision on the MP, they gave limited information on factors that may have influenced the classes' decision.

In the course of discussing the MP's main point, interviewees did discuss four objectives of the lesson: (a) to introduce students to the MP, and invite them to develop theories that might account for it, (b) to encourage students to evaluate proposed theories for the MP, (c) to interpret historical investigations to test students' theories, and (d) to recognize interpretation of results may be ambiguous, but this does not preclude consensus. In summary, the interviewees focused on what was happening with the moths and the theories that would best explain the phenomenon. In addition they revealed in part, how they made their choice of explanations. They also revealed a tension between the choice of explanations and their interpretation of what the stated mystery was. As mentioned above, about half of the interviewees thought the mystery was to discover why the light moths were turning dark. The tension manifests itself in the interviewees' inability to give acceptable explanations about the scientific concepts discussed next.

Scientific concepts. The interview questions about variation and natural selection were broad and open-ended, this resulted in responses that varied in depth and focus. To determine what the interviewee understood about the concept each of the four questions was coded separately and then together. The same concept was often discussed in responses to different questions and at times the information was contradictory. Therefore all four questions about variation and natural selection were considered when deciding if a scientifically acceptable answer was given. Responses that contained for correct and incorrect information were considered hybrids and not acceptable overall. Examples of hybrid responses are included below for the concept of variation, which is the first scientific concept discussed by the interviewees.

Variation. The first interview question about variation asked “What is variation?” An acceptable response would be a definition of the concept, which would include the idea that individuals within a species have the same traits or characteristics, but these vary from individual to individual. Examples could include hair, eye or skin color or the different colors of peppered moths. The companion question “How does variation happen? Is asking for an explanation of the origin of variation. The response should include random mutations and/or sexual selection. The answer may not include those exact terms, but may talk about changes in DNA that occurred randomly, or an individual receiving genetic material from both parents.

For the concept of variation most of interviewees in both groups gave a response, and majority of the responses (~ 80%) were acceptable (Table 23), For example, F5 stated “different traits or characteristics that the moths had…they’re the same species, but they’re different colors.” This statement is acceptable. The interviewees that did not give

an acceptable answer all had hybrid answers. For example, interviewee F3 stated that variation was “different traits…like eye colors, different skin colors…” This description of variation was fine, and if it was by itself would have been acceptable. However F3 also stated that variation happened between different species, so the answer was not acceptable. All the hybrid answers confused the concept of a species, either stating variation happens between species, or using different species as examples of variation within the same species, such as the finches on the Galapagos Islands (Table 24). These alternative conceptions are different from those chosen by the participants on the CINS. The distractors for the variation questions were limited to the idea that all individuals are the same, or that variations have no influence on survival; therefore all the participants who had an incorrect answer chose one of those conceptions. The interviewees with incorrect answers made no mention of individuals being identical or nearly identical. Most interviewees in both groups also answered the question: “How does variation happen?” but they did not have much success with this question.

When answering this question they talked about two concepts, the origins of variation and how variation is inherited. The majority of interviewees (~ 90 %) answered the question with reference to the *origin of variation*, which was my original intention. The interviewees in both groups had a difficult time with this concept. Only one interviewee gave an acceptable explanation in the control group, and five in the treatment group. A few things emerged the incorrect answers. First, four interviewees in the control group and two in the treatment group talked about the need for organisms to change in response to the environment or in order to survive, and this need directed the individual’s change. This use of intentional change is not surprising, as discussed in the CINS results,

intentional change was the most frequently used alternative conception for origin of variation (Table 18). The literature also supports the students' appeal to intentional change when formulating evolutionary explanations (Evans, 2008).

Table 23

Interviewee Response According to Concept

Concept	Response type	control n=15	treatment n=14
Variation	mentioned	11	13
	accepted	9	11
	not accepted	2[a]	2[a]
Origin of variation	mentioned	13	13
	accepted	1	5
	not accepted	12 (3[a])	7
Variation Inherited	mentioned	6	9
	accepted	5	9
	not accepted	1	0
differential survival	mentioned	14	12
	accepted	5	7
	not accepted	8 (2[a])	5
change in population	mentioned	4	3
	accepted	4	2
	not accepted	0	1

Note[a] hybrid answer: part of the answer had a statement that was correct, while another statement was incorrect.

Another factor that may have influenced the interviewees' inability to give an acceptable explanation about the origin of variations is their belief that the dark moths have always existed. For example, in answer to the question "How does variation happen?" F5 from the control group, said the traits had always been in the individual and natural selection "pulled" the favored trait out for the particular situation. Another interviewee, F10 answered that the black moths had always been there and just their numbers changed. In the treatment group, two interviewees, S7 and S12 also said variations have always existed and changing circumstances changed their numbers. S12

Table 24

The Alternative Conceptions Given in the Interview Responses

Concept	Alternative Conception
Variation	Variation refers to differences between species. • F3 "Uh, **like different traits between species** or anything really, just traits between two different things." Different species are considered the same species • S6 " …the Galapagos Islands and the birds…how their beaks were changed and different,…some of them had small beaks, some of them had large beaks, **like why they are considered to be the same bird, but then they had different variations.**
Origin of variation	Individuals choose to change • F3 "…and the ones that don't have…well the brightest feathers **will try and somehow make themselves brighter** in order to attract more female mates." Two species mate to produce new traits • F13 "…but I meant like **two different species coming together and mating, successfully mating** is what I was referring to." Trait was always present & environment influences change • F5 "…it's like **their traits had just always been within them and natural selection pulled like, just favored one of the traits**, like depending like where they are, **to make them more suitable for where that location is**.." New environments cause change • S13 "…it occurs because…if you have something that's taken out of its environment, **and placed in a new environment, it needs to adapt** so that it can become better suited for its new situation…"
Differential survival	Individuals want to survive so they learn to adapt to the environment • F3 "…(moths) they get darker through natural selection…**if they want to live they have to learn to adapt to their environment."** "…they have to um, I guess select **which kind of type they want to be**." Survival of fittest refers to individual attributes • S6 " what bluegill…is most fit based off this chart, and there was like different characteristics, … **how fast it can travel.** Survival refers to individuals in new environments • F4 " **A house dog were to run away**…**if they don't adjust to wild, and the surroundings, their new surroundings they probably won't last very long**, it will eventually die out." Survival of fittest refers to having the most mates • S6 "**maybe the amount of times it mated**," Competition for survival is between two different species • S2 "Species with most resources and the **most mates**…can most easily adapt to their environment…**so they'll survive most out of the other species**"
Limited survival	Competition for survival is between two different species • S1 "…**the species that** don't have enough food or **can't compete, they'll die off and another species will come take its place or survive instead."**
Limited resources	Competition for limited resources • S1 "…**the species that don't have enough food** or can't compete**, they'll die off and another species will come take its place or survive instead."**
Speciation	Two species mate to produce new species • F14 "…**one species of finch or whatever, mated with another** they could **mutate and form a new species**, or change the genetics in some way."

said "…like what we learned with the moths…the darker moth just didn't pop out of nowhere, like it was always there," While only five interviewees explicitly talked about the trait or the black moths always being present in conjunction with the origin of traits, 10 in all talked about it in other questions. Most said that the lesson or their instructor stated that the dark moths had always existed. The lesson's script does not talk about the dark moths always existing, but the class videos did reveal all of the instructors mentioned that a small population of dark moths did exist at the time of the industrial revolution, this may be where the interviewees' belief about the dark moths always existing originated.

Finally, two interviewees in the control group, F13 and F14 stated that variation happened when two different species mated new variations happened in the offspring. The remaining incorrect answers did not have a specific alternative conception or factor that could account for them, instead the answers were just unclear.

Interviewees also talked about the concept of *variation inherited*. About 40% of the control group and 60% of the treatment group talked about this concept. Less interviewees discussed variation inherited versus origin of variation; which is probably related to how to they interpreted the interview question. All but one interviewee in the control group gave an acceptable answer. It is apparent that the interviewees' who chose to talk about this concept understood it. This concept was also not very difficult for the participants on the CINS instrument with ~ 70% answering the questions correctly in both groups. The next two interview questions were on natural selection.

Natural selection. The first interview question on natural selection was "What is natural selection?" and the second was "How does natural selection work?" The concepts discussed by the interviewees for both of these questions were *differential survival* and/or *change in population*. Acceptable answers for the concept of *differential survival* could include statements about individuals who have traits that help them survive, reproduce and pass down their genetic material to their offspring. Answers may talk about the offspring surviving to an age where they can reproduce. Answers would not have to include complete descriptions of different survival, but what was discussed had to be correct. For example, S4 in the treatment group stated "…like in the moths how the darker ones were better camouflaged, so because they were less easily seen by predators, then they were able to survive better, and were able to reproduce and then pass it on.". Answers for the concept of *change in population* should talk about the unequal ability of individuals to survive and reproduce will lead to an increase of frequency of certain traits in the population. For example, F11 in the control group said "…the ones that are more fit to survive are more likely to reproduce, and so over time the population will change because there's more of a certain type."

The concept of *differential survival* was difficult for the interviewees. About 90% of both groups talked about differential survival in connection to natural selection; only about one third of the control group and one half of the treatment group had acceptable answers. Interviewees used a variety of alternative conceptions in their explanations about natural selection and differential survival. For example, F3 in the control group talked about peacocks intentionally making themselves brighter in order to attract mates. F3 is first talking about intentional change which is an alternative conception about the

origin of variation. Then F3 discusses the ability of attracting mates as a factor of fitness, which is an alternative conception about differential survival. Another interviewee, F1, stated the moths saw other moths being eaten, so the adapted because they did not want to get eaten. This explanation is also about intentional change. The reason is about personal survival and not about passing on traits through reproduction.

Other interviewees in the treatment group, S1 and S2 talked about fitness as one species competing with another species for resources such as food, mates and ultimately for survival (Table 24). These examples show that the interviewees did not just hold alternative conceptions directly about differential survival, but ones about the origin of variation, limited survival and other concepts. Not all the incorrect answers expressed explicit alternative conceptions; some were either confusing or unintelligible. The distractors on the CINS for differential survival were two alternative conceptions about fitness so all the participants' incorrect answers on the CINS reflected this. The alternative conceptions were that fitness is equated with an individual's traits such as strength, speed or intelligence or longevity, or it is equated with the number of mates an individual has. The interview responses show that explanations can be complex and may include more than one alternative conception. As discussed in question RQ1, differential survival was one of the most difficult concepts on the CINS and as stated earlier in this section, for the interviewees.

Another concept related to natural selection is *change in population* and a few interviewees in both groups talked about it. All but one response were acceptable, so for the interviewees that mentioned this concept most of them understood it. The incorrect response was confusing, but did not contain any explicit alternative conceptions. The

participants on the CINS initially found this concept to be problematic. The pre-test scores ranked this concept as the third most difficult. After instruction both groups had at least 60% of participants answer both questions correctly for this concept.

As discussed above, the interview responses for the questions involving variation and natural selection did reveal alternative conceptions and they are summarized in Table 24. When comparing the type of alternative conceptions used by the interviewees versus by the participants on the CINS, there were both similarities and differences. There were similar alternative conceptions used for the concepts of *origin of variation* and *differential survival*. For *origin of variation*, the idea that individuals can change intentionally. For *differential survival,* fitness was referred to as individual attributes such as strength, speed intelligence or longevity, or the ability to attract multiple mates. There were three different types of alternative conceptions used only by the interviewees First, interviewees used alternative conceptions about species in five concepts: (a) *variation*, (b) *origin of variation*, (c) *limited survival*, (d) *limited resources*, and (e) *speciation*. All six of the alternative conceptions on species involved different species, such as two species mating to produce a new species, or two species being referred to as the same species (Table 24). Second, the interviewees talked about change occurring or needing to occur because an individual was placed into a new environment. This alternative conception was found in *origin of variation* and *differential survival.* Third, some interviewees thought traits are always present in an individual and the environment causes it to express itself. This alternative conception was found in *origin of variation*.

When the alternative conceptions used by the participants in the interviews are considered in reference to the four explicitly targeted ones embedded into the MP lesson,

only one was part of the responses. For the concept of *origin of variation*, the Lamarckian idea of intentional change was used by participants in both groups. Twelve participants in the control group gave incorrect responses and five of those referred to intentional change, seven participants in the treatment group gave incorrect response and two of those were about intentional change. It is interesting that for the concept of *origin of variation*, this misconception, which is explicitly discussed in the MP lesson, was the one used the least in both groups. Also, the treatment group used this misconception less than the control group, which supports the idea that the story version of the MP lesson has advantages over the original version.

In summary, several things emerged from this portion of the question. First, the interviewees in both groups had a good grasp of the lesson's main point, which was to solve the mystery about the dark moths by finding the best explanation for the phenomenon. They had more difficulty articulating what the actual mystery was. About half of each group thought the mystery was to explain how the light moths were turning dark. The majority of the interviewees chose natural selection as the best solution to the mystery. In addition, most of them felt the class came to the same conclusion. Second, after instruction the interviewees in both groups had difficulty with the concepts of *origin of variation* and *differential survival* which was not unexpected because the CINS results demonstrated the same issue. So in spite of the interviewees choosing the natural selection as the preferred explanation, they had a difficult explaining that process. While both groups had difficulty with the above concepts, Table 23 displays that the treatment group had a higher percentage of correct responses for these concepts. For the concept of *origin of variation*, which is the most problematic concept for the participants, the control

group had a more difficult time with only one of 13 correct responses. The treatment group had five of 13 correct responses. So while these concepts remain challenging, the use of the story version of the MP lesson appears to have a positive impact on these alternative conceptions.

Third, the responses revealed that the interviewees hold alternative conceptions outside of those included in the CINS instrument, and that they sometimes gave responses with both correct and incorrect statements. This shows that the explanations constructed by the interviewees are complex and sometimes contain alternative conceptions about a variety of concepts. The interviewees' use of both correct and incorrect statements in their explanations supports how tenacious alternative conceptions are to instruction. In addition, the prevalence of incorrect statements about species was unexpected. These statements were found in five different concepts which suggests that issue may be wide spread. It may be beneficial to review the concept of a species with students periodically during the course.

Research Question 3

RQ3. What are the similarities and differences in participants' experiences, as revealed in the interviews, in both the traditional historical and the story?

While the interview questions discussed above were interested in the participants' explanations of variation and natural selection, the remaining questions were interested in the participants' experiences with the MP lesson. There were two foci to this group of interview questions. First, questions IQ5a, b focused on the MP as a lesson, and asked the participants to talk about their thoughts of the lesson as a whole, its components, and the perceived benefits and disadvantages. In addition, the treatment group was asked

question I8a-d about the lesson's PowerPoints, because they had been modified from the control group. The interviewees were encouraged to talk about their positive and negative impressions and share what they might change and why. Second, questions IQ6a-c and IQ7a-c focused on the idea of a lesson presented as a mystery. Interviewees were encouraged to talk about other lessons in the class they felt were mysteries and what made the similar or different from the MP lesson. They were also asked about the advantages and disadvantages of mysteries in the classroom.

MP lesson. Both control and treatment groups provided feedback the MP lesson. This feedback was both positive and negative and some participants gave suggestions on how to improve the lesson. The majority of participants in both groups gave feedback on the lesson. The comments fit into two broad themes. First, the interviewees talked about different aspects of the MP lesson. The participants' comments in both groups were focused on three areas: (a) the topic, (b) the lesson format, and (c) the components of the lesson. Second, the interviewees' overall feelings of the MP lesson were apparent from their comments. While many interviewees gave both positive and negative feedback, I was able to use their feedback to categorize their stance toward the lesson.

Topic. The topic of the MP lesson consists of the phenomenon of industrial melanism of the peppered moth and the theories and scientists involved with the phenomenon. Most interviewees in both groups commented on the topic and the majority of the comments were positive (Table 25). Both groups gave similar positive comments of different theories as positive aspects of the topic. Negative comments tended to vary. There was one interviewee (F1) in the control group that thought the topic was repetitive because only moths were used in the lesson. Three interviewees in the treatment group

Table 25

Number of Participants who Shared Positive or Negative Comments

Theme	Control n=15		Treatment n=14	
	positive	negative	positive	negative
Topic	11	1	9	4
Format	11	6	8	6
Component	7	6	3	7

spoke about evolution in a negative light. One did not find the topic interesting (S2), another found it tough and confusing (S5), and one (S12) did not feel the topic made sense in light of their religious views. Finally, in the treatment group, one interviewee did not find the scientist's views relevant (S6).

Lesson format. Most interviewees in both groups gave comments on the MP lesson, which consisted of how the lesson was structured and delivered. While the majority of the comments from both groups were positive, interviewees gave more negative comments on the lesson then they did for the topic (Table 25). General comments characterized the lesson as "good", "enjoyable" and "entertaining" Both groups commented on the lesson being planned and/or delivered well. They also mentioned the use of a mystery as a positive part of the lesson, S10 in the treatment group said the mystery sparked interest and kept the students motivated so they would want to know how the mystery ended. One interviewee in the control group mentioned that the inquiry-based format was good (F8). The treatment group added that soliciting the students' own ideas was good (S1) as was the analyzing of the scientists' theories (S7). One interviewee (S8) said that from the perspective of the teacher the lesson was good, and S3 thought learning evolution by using one species was a good approach. Both

groups stated that the lesson provided the benefits of increased thinking by the students. In addition, F8 in the control group, felt that the lesson would increase learning. Both groups gave similar negative comments. There were comments about the lesson being too long, and there being too much to learn. Both groups negatively commented on comparing the theorists' views of the phenomenon. These comments included that this process was not relevant, and it was boring and repetitive. One interviewee in the control group F2, did not like the fact that the MP had no definite answer, and one interviewee in the treatment group did not like the inquiry-based format (S12).

MP Components. The components referred to by the interviewees were the PowerPoints, worksheets and videos. Approximately half of the interviewees in both groups commented on the lesson components. In the control group, about half the comments were positive, and in the treatment group about one-third were positive. In both groups these comments focused on the PowerPoints and the worksheets. The interviewees thought the PowerPoints were good in general and interesting (F5, S9). Specifically, they felt that the visual images of the moths were helpful (F2, S4). They also felt the worksheets comparing the theorists were helpful.

Negative comments from both groups also included the PowerPoints and worksheets. Two interviewees found the PowerPoints boring (F6) and even "painful" (S14), and one interviewee (S9) thought there was too much time devoted to the worksheets. Other negative comments included the Annenberg video, which was shown on day three of the MP lesson. This video was a break in the normal format and did not involve the MP directly. Both groups had a few interviewees who mentioned this video in a negative light. Complaints included the video was too long (S6), not placed well in the

sequence of the MP lesson (S13), and hard to take notes while watching (S7). The MP lesson also incorporated a video on Darwin's theory of natural selection and common descent and one interviewee in the control group did not like the Darwin video (F11) because they felt it was bias in favor of Darwin. Both groups commented that the lack of hands-on activities was a negative aspect of the MP lesson, and that this was different from other lessons in the course.

For all three areas of the lesson, the topic, the lesson format and the components, both positive and negative comments were shared. Some interviewees provided both types of remarks. However, overall the comments were positive. Interviewees also provided suggestions for improvement. The control group had eight interviewees that shared suggestions, and the treatment group had 12. These suggestions can also be sorted in to the areas of topic, lesson format and components. For the topic, both groups suggested that more information be added to the MP; such more background on the scientists, including their research (F10, S10), and more recent research by scientists (S7). Some interviewees want more examples of variation to go along with the moths (F1, S8).

One of the issues interviewees had about the lesson format was that there was a lot of information in the MP. Suggestions on how to handle it varied from adding more time (F6), to summarizing the information and shortening the lesson time (F12, S6). Another aspect of the MP lesson is that it does not have a definite answer, and while this was not an issue to the majority of interviewees, others found the lack of an answer troubling, so their suggestion was to provide a concrete answer (F2). Other comments

focused on group work, one interviewee asked from more group discussions (S3) and one wanted the worksheets completed as a group (F9).

Interviewees in both groups made several suggestions about the lesson's current components and about what was missing. All the suggestions about the PowerPoints came from the control group. One interviewee (F7) liked the PowerPoints but suggested they be updated because it was "Too nineties", and F1 suggested that the PowerPoints include more visuals or just be remade as a video. The only suggestion about the worksheets came from treatment interviewee S10, who felt there should be more information included on the worksheets. The rest of the suggestions were about additions to the MP lesson. There were eight interviewees between both groups who suggested adding hands-on activities. They felt the lack of hands-on activities made this lesson different from other lessons, and that hands-on would be beneficial. Three interviewees wanted experiments (F9, S12) or something they could observe happening (S8). Two interviewees (F12, S13) thought actual moths may be helpful. S5 suggested some type of artifact, because it helped them visualize information. S5 expressed concern about the exam because of the lack of hands-on activities. S1 suggested that even the use of computers to research the theorists or their own ideas would be helpful. Another addition was suggested by F12, who thought the exploration of the scientists' moth research should be a jigsaw activity so that the student groups would become experts and teach each other. Of all the suggestions the addition of hands-on activities was by far the most common.

In addition to asking both groups in general about the MP lesson, I asked the treatment group a series of questions specifically about the PowerPoints. The rationale to

do so was based on the fact that the treatment group lesson used the modified PowerPoints and scripts, and I wanted to see if they would discuss the PowerPoints as a story, with-out prompting. The interviewees gave more positive comments about the PowerPoints, but the essence of the comments were the same as earlier questions. They focused on descriptors such as “interesting,” relevant” and “understandable” but nothing to suggest that they were thinking of the PowerPoints as a story. The amount and type of negative feedback was similar to earlier questions. The interviewees were asked to talk about how the three PowerPoints were tied together from one day to the next, again in an effort to see if they would talk about the structure of the PowerPoints as a story. The responses did reveal that most of the interviewees felt the PowerPoints and scripts were tied together smoothly and that the recap of the previous day was helpful, but there was no indication that they felt they were being told a story. In summary, this set of questions gave confirmation to how the treatment group felt about certain parts of the MP lesson.

Overall both groups provided similar comments and suggestions, and while most the comments for each area of the lesson were positive it does not reveal how the interviewees individually feel about the MP Lesson. I considered the interviewee responses to questions IQ5a, b from a holistic perspective so I could categorize their stance toward the MP lesson as positive, neutral or negative.

In both groups interviewees shared comments, Table 26 displays the breakdown of these comments into the number of interviewees who made positive, negative or both types of comments. To decide the stance of each interviewee on the MP lesson, I first considered if the interviewee made only positive or negative comments. If this was the

case it was simple to assign their stance as positive or negative. For example, S6 in the control group had this to say about the MP lesson:

> "It was very hard to stay focused, I thought the MP itself was, like dragged, between the three days,…So that was kind of dragged out…, we kept discussing the color of the moths, and we'd go on to the researchers, then we had to keep doing the researcher's point of views. And I just, I don't know if it was maybe just him talking too much or what it was, I just felt it was dragged."

This interviewee had nothing positive to say about the lesson and it was clear their stance was negative. So for the interviewees that only shared positive or negative comments I assigned them either a positive or negative stance, as displayed in Table 26. If an interviewee made both positive and negative comments, I considered the overall tone

Table 26

The Number and Type of Comments Shared by the Control and Treatment Groups

Number of interviewees who shared:	Control N=15	Treatment N=14
comments	15	14
positive comments	14	9
negative comments	9	8
only positive comments	6	4
only negative comments	1	4
both types of comments	8	6
positive stance	12	9
negative stance	3	5

of the response in order to assign a stance for example, S7 said:

> "Yeah, um, Rudge had figured out a way to get us involved and get us working on it, which is like the worksheets and getting us to being able to voice our opinions and then voice which theory we believe the most. So I don't really know if there would be any negatives to the unit, beside there were the points where I was like "I feel like taking a nap" and but other than that it was really good…the one time that comes to mind was, we watched a video and it was of one of the theorists getting the butterflies off the trees.

This interviewee felt positive about the MP lesson as a whole but did acknowledge part of the PowerPoint was boring. Even with this negative comment, it is clear that the stance of this interviewee is positive. This is in contrast to interviewee F1 from the control group had this to share:

> "Oh. I thought it was kind of repetitive. I felt like we learned the same thing every day. ..Just like, you know I keep going back to like that whole moth thing cause I felt like we looked at that like every day.. Like the moth example I'd rather have like a broader range of examples.
>
> In response to the question of why the moths were in the unit the interviewee said:
>
> "…(the moths) I think they were a good example. " That like everyone knows what they are, like they have heard of moths before." (good example) cause I like understand the moth,"

This interviewee thought the moth example for the lesson was good but because it was the only example they thought the entire lesson was repetitive and therefore I assigned F1 a negative stance. There were eight interviewees from the control group and six from the treatment group who shared both positive and negative comments. In the control group, I assigned six of the eight a positive stance. For the treatment group, I assigned five of the six a positive stance. Overall, in the control group 12 interviewees had a positive stance toward the MP lesson and nine did in the treatment group. Therefore the majority of each group had a positive perspective on the MP lesson.

Mysteries. The MP lesson is a mystery. It was designed to introduce a phenomenon in such a way that the students would have to explore different theories and the work of scientists in order to pick the best solution. Interview questions IQ7a-c asked the interviewees to talk about lessons formatted as mysteries. They were asked what were the advantages and disadvantages of having a lesson as a mystery. This question was asked about mysteries in general, because I wanted the interviewees to feel free to talk

about the attributes of mysteries in general, however , the majority of both groups did talk about the MP lesson specifically as a mystery.

The interviewees looked at mysteries from three perspectives: (a) the purpose, (b) attributes, and (c) benefits. First, the majority in both groups talked about the purpose of a mystery as trying to figuring something out and picking the best solution. In case of the MP, the interviewees considered the theorists ideas and their own ideas to find an answer. Being able to explore their own ideas was important to interviewees in both groups. For example, one interviewee valued being able to explore their ideas. F4 stated ", oh mystery phenomenon, I was like oh it's not technically solved, and that means there's a bunch of explanations that it could be, and I get to come up with my own explanation as to why it happened. So I always get a sense of curiosity when someone presents a mystery to me." Another interviewee, F3 liked having other peoples' ideas to compare their own with. So the interviewees appreciated the opportunity to consider their own and others' ideas about the mystery as part of the process of solving it.

Second, the interviewees talked about the attributes of mysteries. Eleven interviewees in each group mentioned inquiry-based learning as an important aspect of mysteries. They considered the process of solving the mystery as inquiry. Interviewee S1 said "I think the advantage is the inquiry part of it, we have to use what we know and our resources to come to a conclusion." Interviewee S4 stated "so I think it's always helpful when you presented like with a mystery or like something you actually like think about and go through the inquiry process, and use critical thinking…" In addition to the inquiry process itself being important, the interviewees connected other attributes of mysteries to the inquiry process. For example, both groups had interviewees who said that mysteries

were interesting, F5 said “It made it more interesting for us, I guess because…like we want to find out, it wasn’t just straight like given to us, it was something that we had to work for.” Fun was another attribute mentioned by interviewees. Treatment group interviewee, S10 equated mysteries with mystery novels which are exciting and fun because people look forward to solving them, and “it keeps things interesting in class…it’s being presented in a way where you are more involved in the process and it makes it more fun.” Another attribute mentioned was that mysteries are applicable to other areas of life. Interviewee F10 stated that it is the process of asking questions that can be personally applied to areas such as ones’ own health. So mysteries involve the process of inquiry, and this process encompasses other attributes. These attributes also contribute to the benefits of mysteries.

The benefits mentioned by the interviewees had two foci: Student behaviors and cognitive benefits. The interviewees talked about mysteries promoting behaviors that involve student engagement. Interviewees said mysteries increase attention and focus (F12, S3), patience (S14), curiosity and intrigue (F4, F12), the desire to solve the mystery (F5) and the desire to learn (F2, S6). The second type of benefits were ones that affected cognitive aspects. Interviewees felt that mysteries would increase thinking in general or critical thinking (F15, S11), learning (F8), understanding (S6) and memory (F2). One interviewee, F7 also felt mysteries would promote creativity. Interviewees talked about the attributes and benefits of mysteries from their own experiences with the MP lesson and from the perspective of a teacher.

Teacher perspective. One surprise that came out of the interviews was how some interviewees took the stance of a teacher as they were talking about mysteries. Three

interviewees in the control group and eight in the treatment talked about the advantages mysteries would have in their classrooms. The advantages stated were the same as above but the explanations of the advantages was focused on teaching not being a student. All but one interviewee in each group mentioned that the inquiry format of a mystery was an advantage. For example, S1 states:

> "I think the advantage there is the inquiry part of it, … with young children, you can't just give them facts and expect them to be to understand something, cause they'll Forget about it, but if you let them explore why things are happening…I think it will help them learn cause they'll come to it on their own and that perspective is going to be really great, and critical thinking later on, that's kind of where it's coming from for us, to use critical thinking and inquiry to find the answers we want,"

Other advantages mentioned by multiple people included an increase in thinking, learning, engagement, attention and the desire to learn. As with the interviewees in general, the interviewees speaking as future teachers equated the idea of a mystery to inquiry learning. Inquiry-based learning is considered by most of the interviewees as a positive aspect of mysteries. This was not the case for one interviewee from the treatment group who thought the inquiry format was not good. S2 said "I don't want to like mazed into the answer. I want the answer. So, but I guess it could be beneficial for people who understand learning like that…I'm not that type." However, S2 did admit that as a teacher mysteries would be more fun for children. In summary, the interviewees felt that mysteries have advantages that make them a desirable tool as students and as teachers. I then asked them about the disadvantages of mysteries in the classroom, of which and they did shared a few.

The main disadvantage of mysteries for both groups involved the possibility of not solving the mystery which could lead to frustration (S10) and arguments (F4). The possibility of not solving a mystery stems from the lack of a definite answers, and the need for students to choose the best answer based on evidence. F2 from the control group said "I guess the problem I've been saying, I've been struggling with, is that, like I feel that we didn't come to an exact conclusion." Not having a definite answer leads to issue of not having a correct answer for assessment purposes. For example F14 said "Some people might get a little upset if they can't figure it out, or if they can't come to a conclusion…they won't study for their test because they can't figure out this lesson" The lack of definite answers could also lead to misconceptions, S3 stated that "I think that leaves a lot of room…for misconceptions, because as you are developing theories, you might think of a wrong theory, and that might be what sticks with the student…" At least half of each group talked about a disadvantage connected to the lack of a definite answer. Besides the issues with a lack of a definite answer the only other disadvantage mention was the extra time needed to prepare and deliver a mystery. Interviewee F8 shared "It takes longer to do in class, he could have just told us about Lamarckian inheritance and natural selection… that would have taken us like half a class, instead it took us three days, alright, and it's a lot harder to plan for I would think."

After sharing the possible disadvantages of a mystery, I asked the interviewees if they felt the advantages of mysteries outweighed the disadvantages of mysteries. In the control group 14 of the 15 interviewees answered this question. Thirteen felt that the advantages of mysteries would certainly outweigh any disadvantage. In the treatment group 11 interviewees answered this question and 9 agreed that the advantages

outweighed the disadvantages. This positive support from both groups is not surprising in light of their positive comments about mysteries and inquiry learning. The interviewees talked about the desired attributes and perceived benefits of mysteries, and in general endorsed their use in the classroom.

Other mysteries. I was interested to see if the interviewees considered any lessons or activities prior to the MP lesson as a mystery. The rationale for doing so was two-fold. First, mysteries may be considered a type of story by some, and I wished to see if students mentioned characteristics of stories when talking about mysteries. Second, I also wanted to see what the interviewees thought were the similarities and differences between the MP lesson and other perceived mysteries, in an attempt to discover commonalities. The majority of the interviewees in the control and treatment groups stated there were other lessons they could consider mysteries in the course. Table 27 summarizes the other mysteries mentioned by the participants.

There were three activities that were mentioned the most in both groups: (a) the silver box (b) Netlogo pond simulation and, (c) Galen & heart circulation. As mentioned above, the majority of both groups expressed that the purpose of a mystery was to figure out how something works and/or why something happens. For example the silver box is a small covered box with sticks protruding out. When one stick is moved the other moves in a particular motion, the interviewees said the purpose was to figure out what might be inside the box that would cause the movement of the sticks (F12, S9). In the Netlogo pond simulation the interviewees said the purpose was to figure out what was happening

in the pond scenario with varying amounts of sunlight, algae and different fish species (F6, S13). . The heart circulation lesson involves Galen's theories of the circulation and

Table 27

Course Activities Considered to be Mysteries by Interviewees

Mystery	Control n=15	Treatment n=14
No other mysteries	1	4
Yes other mysteries	14	10
silver box activity	7	6
Netlogo pond simulation	5	3
Galen &heart circulation	3	4
Brassica plant investigations	2	-
misconceptions project	-	2
pre-assessments	-	2
living or not activity	1	1
Owl pellets	1	-
false limb discussion	1	-

has a sheep heart dissection. Interviewees were less clear on the purpose, for example, S8 felt the heart circulation was a mystery to those that had not learned about it yet, and the purpose was to debunk misconceptions about circulation. F7 also considered this activity from a personal perspective, and said "I enjoy the hands on stuff like when we dissected the sheep heart, and stuff that's exciting to me cause I actually can explore, figure out things myself…" Others talked about exploring Galen's model (S10, S12) of circulation as way to investigate if ideas are scientific, and how the heart actually pumps blood through the body. One interviewee (F15) felt the heart dissection was a mystery, but did not feel there was a main point. With the exception of F15, when the interviewees talked about the purposes of these specific activities it supported their discussion of the MP lesson and mysteries in general.

When I asked the interviewees what they thought were the similarities or differences between these mysteries and the MP lesson they did not give a lot of detail, but they did share a few thoughts. These similarities were mentioned by both groups: (a) the purpose was to figure out how something worked or happened (F4, S12), (b) there was no definite solution (F5, S5) and that best answer had to be chosen using evidence (F10), (c) different theories were presented (F11, S12), and (d) students got to generate and use their own ideas (F7, S9). One difference came up from a few interviewees, and this was how the MP lesson had no hands on and more lecture than the other mysteries (F11, S6). Other differences mentioned focused on specific activities, such as the MP lesson and Galen's heart circulation exploring theories where NetLogo does not (F6).

The interviewees shared how they felt about mysteries as a lesson format. Their comments focused on the purpose, format and benefits of mysteries. They felt that the purpose of the mysteries were to pose a problem or question to the students which had more than one possible answer. The students then had to use their own ideas and/or the ideas of others to choose the best answer. Almost all of interviewees considered the mysteries in this course to be inquiry-based, and this is a preferred method of teaching because of the perceived benefits. These benefits include positive changes in behavior such as increased attention and curiosity, along with the cognitive benefits of increased critical thinking, memory, and understanding. The benefits mentioned by the interviewees have been discussed in the literature as benefits of stories, and this connection will be considered further in Chapter 5. While discussing mysteries, the interviewees described attributes and benefits equated with stories, however, there was no explicit mention of stories. The one exception was in the response of interviewee S10

who talked about mystery novels. S10 was talking about how exciting and fun mysteries are to solve and used the idea of a mystery novel as an example.

This responses of the interviewees revealed that they generally had a positive stance toward the MP lesson and felt that the format of a mystery in a beneficial way to deliver a lesson. This research question did not reveal if the interviewees considered the MP lesson as a story. The final research question considers this.

Research Question 4

RQ4. What do the interviews reveal about the participants' awareness of the story and its narrative elements embedded in the story approach?

This research question focuses on the treatment group because the MP lesson was delivered as a historical story; defined by the 10 narrative elements by (Klassen, 2009) See Table 1 in Chapter One. Interview question IQ9 asked if the PowerPoints could be considered a story. The interviewees were encouraged to give reasons for their answers. This was the last interview question for two reasons. First, I wished to see if the treatment group discussed the PowerPoints as a story which contained particular components. I did not want to bias potential answers by discussing the possibility that the MP lesson was a story earlier in the interview process. Second, an important aspect of the reader/listener narrative element, is that he/she must perceive the account as a story in order to anticipate what comes next. This anticipation is one of the benefits of the story structure (Norris et. al, 2005; Klassen, 2009). Therefore it was important that the interviewees were not prompted about listening to a story during the interview process. This interview question was first explored using emergent coding, and then examined for any description or mention of Klassen's 10 narrative elements (2009).

PowerPoints as story. The interviewees overwhelmingly decided that the PowerPoints could be considered a story. Nine of 14 interviewees thought the PowerPoints were a story. Two interviewees S10 and S3 did say they were experiencing the lesson as a story while it was being presented. For example, S3 said “I kind of thought of it as a story even, just the way it was presented with the kind of the beginning and we explored what was happening and why it was happening and possible, possible endings. “Three more said the PowerPoints could be considered a story, but with conditions. Only two did not think the PowerPoints were a story. All the interviewees talked about a story having a beginning, middle and end or some variation of this. For the interviewees who considered the PowerPoints a story, the beginning-middle-end structure was given as a reason it was a story. This structure was also important for two of the interviewees, S3 and S12, who conditionally considered the PowerPoints a story. They said normally stories have a definite ending, but the MP lesson did not and therefore it was different from most stories, but they still considered it a story. A third interviewee, S11, also considered the PowerPoints conditionally as a story, but said most stories included both major and minor details, while the MP lesson only had major details. The two interviewees that did not consider the PowerPoints as a story also talked about the structure as a factor for their decision. Interviewee, S1, felt the structure of the MP lesson was missing an ending because there was no resolution to the mystery. Therefore, she considered a lack of definite ending as a violation of the structure that would, in her opinion, make the lesson a story. The second interviewee, S14, felt the PowerPoints did have the beginning-middle-end structure, but still did not consider the MP lesson a story. The reasons S14 gave were that the PowerPoints were not related and the lesson was not

interesting enough to compel her to wonder about what was happening next, so the lack of these things outweighed the presence of the structure. In summary, all interviewees considered the beginning-middle-end structure integral to stories, but it in itself is not always sufficient to consider an account a story. The need for a story to have a particular structure that helps the story together is supported in the narrative literature. Altman (p.3, 2008) states that "Stories must be coherent; they must have a distinct beginning, middle and end, they must connect their parts through clearly motivated causes…" Altman's statement supports the reasons S1 and S14 gave for not considering the PowerPoints a story. While the beginning-middle-end structure was important to the interviewees, it was not the only story component mentioned.

Story components. The interviewees mentioned nine other components that they felt belonged in a story. Three of the nine components were mentioned by a least 1/3rd of the treatment group. The next most frequently mentioned component after the beginning-middle-end structure was human agency. Ten interviewees talked about human thoughts, experiences and actions. Examples from the MP lesson included what the researchers were thinking about the moths, observing and collecting moths and experiments with the moths. Six interviewees specifically mentioned characters, while only three of those six talked about background information of those characters. Five interviewees talked about a problem with a resolution being a part of a story. Other components mentioned were plot, main idea, series of events, past time and mystery. While ten components overall were mentioned only two were considered essential by interviewees. These were the beginning-middle-end structure and the problem with resolution component. Only one interviewee talked about the essence of a story being important to the story. Interviewee

S14 felt that a story had to be interesting before in could be considered a story. The interviewees discussed their idea of a story in their own words, and while this provides insight to their thoughts on stories in general and whether they considered the MP lesson as a story, it does not reveal if they were aware of the narrative elements discussed in Klassen (2009).

Narrative elements. As mentioned earlier, IQ9 was considered a second time for the narrative elements. Once this was complete, other interview questions were also reconsidered for any description of the narrative elements. It is important to note that even if instances of narrative elements were mentioned by the interviewees other than question IQ9, it does not imply they understood that they were discussing a story component. I can only make the assumption they were talking about a part of the MP lesson in light of the interview question. Even if the interviewee was not explicitly aware that they were talking about narrative elements, the fact that the narrative elements are mentioned in conjunction with the lesson makes these instances important, because the interviewees were sharing thoughts they felt were pertinent to the MP lesson. Recall from Table 1 in Chapter one that the ten narrative elements were defined by a list of characteristics. For each narrative element, I reproduced the characteristic list from Table 1 and gave an example that represents that characteristic. Some examples covered more than one characteristic, and in these cases the example is found after all of them. All narrative elements and their characteristics were represented in the interview responses, with the exception of the *past time* characteristic of time manipulation. Time manipulation was not written into the MP lesson so this characteristic is not applicable.

Event-tokens.

- Particular occurrences involving particular actors at a particular place and time.

  IQ9 S9 "…we started at the beginning and then we worked out way all the way up to almost present day with different ideas and different things that people have done,…"

- Are chronologically related

- Involve a unified subject and are interconnected

  IQ9 S9 "…they (PowerPoints) were interrelated. I think if they were very separate pieces of information, very chunked out, it wouldn't come across as a story. But because it was one main topic, one main area of focus and it progressed along, that's why you could look at it as a story.

- Later events seen as significant in light of earlier events.

- Lead to changes of state.

  IQ9 S2 "…a lot of people liked to collect the white moths, and then over time and over the air pollution and the factories, and blah, blah, they (the collectors) started noticing that the white butterflies were decreasing and the black butterflies were increasing

Narrator.

- The agent relating a narrative.

  IQ9 S10 " …you know Kate would read some text and talk…like almost the individuals as if they were characters in this story."

- Determines the point and purpose of the story to be told.

- Selects events and the sequence in which they are told.

- Fashions a sequence of events into a significant whole.

  IQ9 S10 " She gave us some background information on them, even their personalities, which um that did make it more like you are reading a story,…you're starting at one point then you end at the end, you know just like a story would,…your know, your beginning, your climax and then your conclusion"

Narrative appetite.

- Desire created in readers and listeners to know what will happen

  IQ8 S10 "I definitely think it sparked interest for people, splitting it up to three different days, was interesting because I was kind of always waiting to know what

was going to happen the next day…I am actually looking forward to finding out what's going on next class."

Interviewer: Can you tell me exactly what is was that sparked your interest?

S10 "…the mystery of it, you know, starting with a question and having to find a solution…"

- Based on a range of possibilities that creates anticipation and suspense

IQ9 S4 "…and now that I think about it the problem, there's a problem and then multiple solutions were presented…"

Past time.

- Narratives concern the past.

IQ9 S7 "…we got different information at different points of time, so we had like a beginning point and an ending point, and there was like timeline, in-between, time past, and over time,…"

- Narrators can manipulate time in relating narratives.

Not Applicable

Structure.

- Narratives start with imbalances, introduce complications, and end in success or failure.

S13 "…it is a story, because you start out with your question and then throughout the time that goes by you have solutions…and things are getting shutdown or supported, and at the end there's kind of the conclusion…"

- Narratives are structured around the sequence of plot events.

IQ9 S12 "…the characters I guess would be the theorists, and the moths themselves, and the plot was just where were they coming from (the moths), the explanation for just the differences in the types of moths and how were they able to increase so much? how did that happen?, I question that would be the plot.

Agency

- Actors cause and experience events in narratives.

IQ9 S5 "…and the character would be obviously the, …three people who came up with these theories and then there were people who actually went out and did the research,…"

- Narratives involve human beings or other moral agents.

IQ9 S7 "…we go inside the theorists' brains, so we go to know them, we got to know how they think and what they believe, so getting to know characters is a big thing of a story."

Purpose

- To help us better understand the natural world and humans' place in it.
- To help us imagine and feel the experiences of others.

IQ7 S3 "…and I think it helps kind of get an overall picture of why things happen, um because we didn't look at just what was happening with the moths, but we looked at what was happening in the environment and in the time period, and the um, what different scientists during it were coming up with, and why did they come up with that? And you know what did the scientists in 1905 and 1960, like what was the difference between their thinking and just all that. (ok) stuff, really, really is good."

Reader

- The reader must interpret the text as a narrative in order to approach it with appropriate expectations and anticipations.

IQ9 S10 "I did think of it as a story, as it was presented to us,…"

Effect of untold

- Missing information increases curiosity.

IQ7 S3 "I kind of liked looking back and tying in things that had happened in the past, and you know why, why did this happen? And do things like this still happen today? And that kind of stuff and as a future teacher I think using kind of mystery or um, mystery units like this gets the attentions and holds the attention of students …"

Irony

- The narrative events are in opposition to the reader's expectations.

IQ9 S7 "Yeah, (ok) I mean there was like blurry point at the end but,"
Interviewer. [laughs], The blurry point at the end?
S7 "Yeah, where they just threw it at us, but,"
Interviewer. The one was?
S7 "The butterfly that."
Interviewer. The middle, the middle-colored butterfly?
S7 "Yea, that threw us all off, we were like "What?" I just remember [laughs] it was crazy, but that confused a lot of people, but, I don't know, I feel like, it just, I guess when I think about it, I feels a lot more like a story than when I was learning about it in class."

Three interview questions provided all the examples of narrative elements, IQ7 which asked about mysteries, IQ8 which asked about the modified PowerPoints and IQ9 which asked about stories. Most of the examples came from IQ9, which is not unexpected. Question IQ7 provided examples of the *effect of the untold* and *purpose* and IQ8 provided an example of *narrative appetite*. It is encouraging that some of the interviewees related to the lesson in a manner that made it possible for them to describe characteristics of all ten narrative elements. It is especially encouraging in light of the fact that none of the narrative elements were explicitly mentioned by the interviewer, so their inclusion was driven by the interviewees. This may suggest that at some level they were aware of these elements in the MP lesson.

Narrativity. The concept of narrativity was described by Norris et al. (2005) as the extent to which a narrative includes all the narrative elements, of which they discussed eight. Norris and his colleagues wondered if some of the narrative elements were more important than others. They proposed that event-tokens, past time and agency were of primary importance and the remaining five may not be necessary for the account to be considered a narrative. Ten of the 14 interviewees talked about human agency and gave examples of the scientists conducting research; this does support Norris et al. suggestion that agency may be of primary importance. The interview responses did not point to event-tokens or past time as being of primary importance, they did suggest that the presence of other narrative elements were important.

One difference that emerged was with the narrative element of narrative appetite. Norris et. al suggests that the lack of narrative appetite could mean there is a poorly told story, not that there is not a story. Interviewee S14 said the MP lesson was not a story

because it did not have narrative appetite. S14 said a story “it is interesting, kind of grabs your attention, like it’s real easy to be engaged, kind of like wondering what’s going to happen next.” S14 acknowledged there was a beginning middle and end structure to the MP lesson, but this was not enough for it to be a story. It may be that stories must contain specific components to be considered stories and if a component is missing its absence is noticed by the reader/listener. Another difference was with the narrative element of structure. The interview responses show that the structure of a story was important to the interviewees. They said that the beginning middle and end structure was a part of a story and it was the only component mentioned by all fourteen interviewees. In addition, the presence of a solution or ending was important to the interviewees. As mentioned earlier, two interviewees felt the MP lesson was a story but the lack of a definite ending and one interviewee did not classify the lesson as a story because of the lack of an ending. In summary, the interviewees as a group, explicitly talked about the narrative elements of agency and structure. Individual interviewees talked about narrative appetite being important to stories, and all the narratives elements were at least described by one interviewee. Two of the narrative elements, effect of the untold and irony, were not in Norris et al., 2005, but were discussed in Klassen (2009). Klassen did not expand on the idea of narrativity, but he did say science stories should contain all ten narrative elements, which of course the MP lesson does. The interview responses for question IQ5 provide support that the majority of the interviewees not only considered the MP lesson a story, they were aware of the narrative elements embedded in the story.

As discussed above, the focus of this study was to empirically test the efficacy of a science story constructed using Klassen’s (2009) narrative elements. This was

accomplished by evaluating the participants' learning outcomes and the experiences in reference to the MP lesson. The results are focused in two main areas, the participants' explanations with regard to questions about natural selection, which included both scientifically correct conceptions and misconceptions, and the participants' responses about the MP lesson.

The participants' explanations were documented through the CINS responses, and the interview responses. The treatment group's CINS post-test results demonstrated significant gains in both average scores and explanatory coherence. This means that after instruction the participants used more scientifically correct explanations. In the treatment group there was also a significant decline in two of the misconceptions explicitly discussed in the MP lesson. While there was a decrease in the use of misconceptions, there was still an insufficient number of participants who gave correct explanations. The most frequently given misconceptions in this study were consistent with previous research, which included misconceptions about Darwin's concept of fitness and Lamarckian ideas of origin of variation and population change. The interviews revealed that some explanations contained both correct and incorrect conceptions, suggesting the participants did not have a complete understanding of these particular concepts. In addition, a noteworthy amount of the participants exhibited misconceptions of the concept of a species. While some of the misconceptions revealed in the interviews were the same as those in the CINS responses, the interviews displayed a wider variety of misconceptions. This is not unexpected because the interview questions were open-ended, which gave the participants a greater opportunity to share their views.

The participants shared their feelings about the MP as a lesson. The responses focused on three areas. First, the participants talked about the MP lesson's components and the lesson as a whole. While individual participants had issues with specific components such as the PowerPoints or the worksheets, no one component was problematic for a majority of the participants. The participants did call for more hands-on components to be added to the lesson. The majority of the participants had a positive stance to the MP lesson. Second, the MP lesson was designed as a mystery in which the students explored three scientific theories that attempted to explain the mystery phenomenon. This is an inquiry-based lesson and the students chose an explanation based on their interpretation of the available evidence. As with the lesson in general, some participants did not appreciate the inquiry format, but the majority of the participants did. They not only enjoyed the inquiry format as current students, they expressed appreciation as future teachers. The participants felt mysteries increase student engagement and have positive impacts on critical thinking, learning and memory. Third, the final interview question asked the treatment group about the MP lesson as a story. Narrative theory states that it is important for participants to perceive a story as a story for the positive impacts to be felt, and therefore it was encouraging to see that most interviewees perceived the MP lesson as a story. In theory it is the narrative elements that distinguish stories from other forms of discourse, therefore it was also important to see if the participants were aware of any of the narrative elements embedded in the story. The participant responses did contain descriptive examples of all 10 narrative elements, which suggests that they were at the very least implicitly aware of these elements. In summary the story version of the MP lesson did demonstrate learning gains, and the participants held positive attitudes

about the lesson in general and as a mystery. In addition, the participants did feel the MP lesson was a story and were able to describe the narrative elements that make it a story. The next section discusses the conclusions, limitations and implications of the research results.

CHAPTER V

CONCLUSION

This study's essence was about stories and their use in science education, and this final chapter brings the previous ones together to consider the study's key findings. To briefly review, Chapter One discusses how evolutionary biology is a historical science and as such looks to the past to explain the present. Because of this, explanations in historical sciences often employ historical narratives. These narratives are distinct from narratives used in other contexts such as stories, however, these two types of narratives have similar structures which suggests that stories based in the HOS have a role in the teaching of evolutionary biology. Chapter Two explores the use of HOS in science education, including how HOS is incorporated into curricular materials and the research on the use of HOS on learning. Chapter Two continues by looking at the perceived benefits of stories from the cognitive sciences and science education. The chapter considers the theoretical aspects of the structure of stories and culminates with Klassen's innovative work, which was built on previous research from multiple disciplines on the structure of narratives. His work on the construction of stories and narrative elements provides the foundation for this study. Chapter Three discusses how I tested Klassen's (2009) approach empirically, which includes employing two versions of the MP lesson as the control and treatment group's interventions. The episode was chosen for study because it incorporates past scientists' theories and investigations on IM as a strategy to mitigate misconceptions. Chapter Three discusses methodologies used to evaluate the participants' CINS responses, and interviewees' use of alternative explanations about concepts related to natural selection. In addition, this chapter explains how the interviewees' experiences with, and attitudes about the MP lesson were conveyed and

interpreted. Chapter Four reports and discusses the outcomes of this research. Chapter Five discusses the how this study fills the gap in the research discussed in Chapter Two, and the significance of the findings in reference to the story version of MP lesson. This includes learning outcomes, mitigation of misconceptions and the attitudes of the participants on both approaches of the MP lesson. This chapter also considers the limitations and implications of those findings.

Research Gaps

Benefits of stories. As discussed in Chapter Two, educators are interested in the use of stories because of their perceived benefits. These benefits include positive student behavior such as enhanced student engagement (Klassen, 2009) and cognitive factors such as increased understanding and memory (Norris et al., 2005). It spite of the large amount of positive literature on the use of stories, there has been a lack of empirical studies to test of the effectiveness of stories in science education, specifically in the topic of evolution. This study was intended to begin to address the lack of empirical studies about the use of stories in science education.

One focus of this study was to address the gaps in Klassen's research on the use of a door-opener story on Louis Slotin for teaching physics students about radiation safety (2009). Door-openers are designed to create curiosity in students which should increase student engagement by providing reasons to want to know more. Klassen provided an elegant way to construct a story using 10 narrative elements, and he demonstrated this with the Louis Slotin story. His analysis of the story involved both critiquing the story to assure it contained all the appropriate elements and evaluating

student responses to the story (Klassen, 2009). It is this second portion of the analysis that has limitations. Because the story was a door-opener, the data collected was limited to student generated questions which should have indicated curiosity about the story. Klassen (2009) relied on the fact that most of the student questions generated were "how" and "why" questions to support the idea that the Louis Slotin story raised curiosity. Curiosity and wanting to know more are important to support the presence of narrative appetite, and therefore interviews should have been conducted to support the assumption the "how" and "why" questions were due to curiosity, and not in response to an assignment. The lack of interviews was a missed opportunity to explore the presence of narrative appetite and other narrative elements.

My study moved beyond this limitation and conducted interviews to see if the interviewees described the presence of the 10 narrative elements. It is important to remember that the majority of the treatment group did consider the Mystery Phenomenon a story, and they discussed different aspects of stories including the narrative elements. The awareness of the narrative elements lend support for the interviewees' comments that stories, such as the Mystery Phenomenon, have positive impacts on student attitudes and motivation. For example, narrative appetite should encourage students to what to know what is happening next, in other words, students should show curiosity or engagement. Therefore when interviewees either expressed enthusiasm, curiosity, or the desire to know more themselves, or stated that the MP lesson as a mystery would create those reaction in others, they in essence were supporting the presence and effect of narrative appetite. It was encouraging to see that this group not only considered the MP lesson's PowerPoints a story, they described all of the structural components Klassen (2009)

thought were important for the construction of an effective story**.** In addition to the lack of participant interviews, Klassen (2009) stated that he was not able to test for the existence of a narrative effect.

Narrative effect. According to Norris et al., 2005, narrative effect is the positive impact narratives have on understanding, interest and memory. Norris et al., 2005 did state that there was moderate support in the literature for the presence of narrative effect, but they acknowledge that other factors besides the type of text may affect learning outcomes. They state that "to establish confidently a narrative effect" (Norris et al. p. 553), an experiment would have to control for all other text variables, including readers the same, which they say is impossible. This may be overstated as they provide evidence in the literature for narrative effect, and they claim this support is moderate. A research study on narrative effect does not have to be experimental to provide support of its existence. This study was quasi-experimental with a control and treatment group. Chapter Three discusses how the potential bias and differences in student populations were controlled for. While no one research project can definitively determine the existence of a narrative effect, it can add to a growing body of research that supports it. This study provides two lines of evidence that support the existence of a narrative effect. First, the treatment group's CINS results demonstrate significant gains from pre to post-test scores and compared to the control group. This means that the treatment group as a whole had a greater understanding of the concepts. In addition, interviewee (S4) shared that the MP lesson was understandable which adds support to the CINS results.

Interview responses also support the presence of a narrative effect. As mentioned in Chapter Four question RQ4, interviewees in both groups talked about the MP lesson as

being interesting or increasing interest in the material. Statements such as these support the existence of a narrative effect. Another purported effect of narratives is increased memory. While interviewees mentioned that one benefit of the MP lesson as a mystery is increased memory, the design of this study did not include evaluating retention of information, so any effect on memory was not measured.

MP lesson. The previous section discussed how this study's interview responses revealed the presence of narrative elements and gave support for the presence of a narrative effect. In addition to exploring the structural aspects of the story approach of the MP lesson, the interviews revealed how the interviewees felt about both versions of the lesson. The majority of both groups had positive attitudes to their version of the MP lesson, and these positive attitudes were generally focused on the MP lesson as a mystery. These interviewees viewed the Mystery Phenomenon as an inquiry-based lesson, and therefore has many perceived benefits for students. These benefits include improved student engagement through increased attention, patience and curiosity. The benefits also included cognitive advantages such as improvements in critical thinking, understanding and memory. It is important to note that both groups expressed these positive attitudes, no doubt because both versions were mysteries. The mystery gave the interviewees an opportunity to explore the views of past scientists, some of which turned out to be known misconceptions, and to apply evidence-based thinking in an effort to choose the best solution. The interviewees personally felt this process was beneficial to them and would be beneficial as future teachers. It is noteworthy that while both groups gave positive responses about the MP lesson, the treatment group performed better on the CINS and experienced greater declines in misconceptions. So while both versions elicited positive

responses, the efficacy of the story version was superior to the traditional historical version. This points out the final limitation of Klassen's (2009) study.

In the Klassen (2009) study, there was no attempt to evaluate learning outcomes in terms of content gains or in the decline of misconceptions. As discussed above in research question RQ1, the treatment group demonstrated statistically significant changes in average CINS scores and in coherence, which the control group did not. In addition, the treatment group had a large portion (26%) of individuals move from a failing score on the pre-test to a passing score on the post-test. The control group had no net movement out of the failing category from pre to post-test. Research question RQ2 talked about the four misconceptions explicitly discussed in both versions of the MP lesson that were based on past scientists ideas. The results demonstrated that the treatment group had larger declines in those misconceptions than the control groups. The results from research questions RQ1 and RQ2 demonstrate that the participants in the treatment group out preformed those in control group. It is noteworthy that both groups used a historically-based lesson, and the only difference was that treatment group was exposed to the story approach. Lessons based on the HOS are contextual, which is considered beneficial for learning, however, stories are also contextual, and the results showed that in this case, the group that had the lesson with a historically-based story generated better learning outcomes than just the HOS-based lesson.

As Chapter Two discussed, there has been a great deal of interest and use of stories, especially historically- based ones in science education but in earlier interventions it was difficult to determine what was actually meant by the term story. Klassen (2009) provided a way to standardize the construction of a story which opened the door for

testing the efficacy of stories using an empirical study. This current study has confirmed that it is possible to use two versions of the same lesson, one with Klassen's (2009) narrative elements, and one without, to test the efficacy of those approaches. This is important because with the ability to test two different approaches, not only can consistency be established within a study, it can be established between studies, which is needed if research on stories is to move forward.

Narrative explanation. Norris et al., 2005 discussed the idea narrative explanations, which they state must have the characteristics of narratives and explanations. Explanations must make unintelligible concepts clear to students and provide them with a causal account of the phenomenon. The narrative aspects of the explanation must include the narrative elements. Norris and his colleagues (2005) considered two examples, one with agency and one without. They did analyze the accounts for the presence of the eight narrative elements they proposed, and suggested what impacts the explanations may have. Because the paper was theoretical, they did not go any further in the analysis. This meant that the possible advantages of narrative explanations were not tested. My study used a historical story to convey the MP. The results demonstrated that my story is a narrative explanation. First, it contains all the narrative elements, and that there were learning gains in the treatment group. Norris et. al, (2005) discusses that narrative explanations or explanatory stories have a role in science because they are well-suited to explain phenomenon that involve a unique set of events. As discussed in Chapter One, evolutionary explanations are series of particular causal events, and therefore explanatory stories are a natural fit for evolution education.

Explanatory coherence. Explanatory coherence was considered in Kampourakis and Zogza (2009) research on evolutionary explanations. As discussed in Chapter Two, they felt it was not only important for students to give scientific explanations to a variety of questions, they should be able to use the same type of explanation for a number of tasks that require that specific type of answer. The ability to do so is called explanatory coherence. Kampourakis and Zogza calculated explanatory coherence for all five tasks in their assessment, which covered two topics, common descent and natural selection. As I discussed in Chapter Four question RQ1, the Kampourakis and Zogza approach was not appropriate for my study because the CINS instrument covered 10 concepts. It is probable that most assessments will have more than five assessment items and more than two concepts, therefore my method of calculating explanatory coherence has the potential of being using useful for a wider range of instruments. Instead of a student either obtaining explanatory coherence or not, their responses are evaluated by each concept, and in this way the student can receive a score. Instructors can calculate the students' pre and post-instruction scores in order to: (a) determine gains in explanatory coherence, and (b) what concepts remain difficult for the students. Kampourakis and Zogza urged researchers of evolution education, and science education in general, to determine explanatory coherence because it is an important parameter. They explain that students who achieve explanatory coherence have accommodated the new scientific concepts. My study not only follows the recommendation of Kampourakis and Zogza by evaluating explanatory coherence, it provides a method that can be used on assessments with multiple concepts. While my study did address the research gaps discussed above, there

were limitations associated with this study. These limitations are discussed in the next section.

Limitations

The limitations to this study originated from the design and execution of the study and from the research community in general. Both the quantitative and qualitative portions of this mixed-methods study had limitations. The limitations in the quantitative portion involved the CINS instrument. The limitations in the qualitative portion involved the interview process. In addition, the differences in the type of responses in the quantitative and qualitative portions created another limitation in the integration of data.

Quantitative. One limitation involved the CINS instrument. This limitation became apparent when the a priori coding of the distractors was taking place. The instrument was designed to feature commonly held alternative conceptions that are tenacious. There are two questions for each of the 10 natural selection concepts, so there could be six different alternative conceptions for each concept, but instead there are two or three per concept. This is a missed opportunity for instructors and researchers to have at their disposal a larger pool of alternative conceptions for the students to choose from. With more distractor choices, it is possible that other alternative conceptions held students may be revealed. If students harbor alternative conceptions not currently on the instrument they may not be addressed by the instructor.

Qualitative. The qualitative portion of the study involved interviews that asked the interviewees about science content and about their experiences with the MP lesson. The interview questions on the science content were both explanatory and exploratory.

The interview questions on the MP lesson were all exploratory. There were a couple of limitations due to the interview protocols and my own interview style for both the control and treatment group. The main issues were with the questions pertaining to stories and to a lesser extent, those involving mysteries.

First, I did not ask the control group interviewees question IQ9 about the PowerPoints being considered a story. My rationale for not including it this question was because the control group did not have a story presented to them. However, the question responses from the treatment group revealed valuable insight to their perception of a story and it would have been useful to see if the control group shared the same conceptions about stories in general. It was have also been interesting to see if any control group interviewees felt their PowerPoints were representing a story, and why they felt that way. This added information may have revealed more information about what narrative elements were necessary in the construction of a story. In both groups, I should have included more questions about stories in general and the narrative elements. Again, more information about what exactly makes up a story in the interviewees mind would have been useful. When considering the interview responses about mysteries, I realized I should have asked another question about the advantages of stories, much in the same way I did for the mysteries. This may have provided useful information for comparison of mysteries and stories. For example, in the interviewees remarks about mysteries they talked about the impact mysteries could have on student behavior, such as increased attention and curiosity. They also talked about mysteries promoting cognitive factors such as understanding and memory. As discussed earlier all these benefits are also associated with stories, and knowing the similarities and differences between the two

would be helpful. In spite of these questions not being asked, the interviewees in the treatment group revealed useful information about stories and the narrative elements. Implications of these limitations are discussed in the next section.

The interviews also had limitations due to my interviewing style, and this was revealed during transcription process. I noticed two things about my style of questioning while transcribing the interviews. First, I noticed that at times, I ran over the interviewees' turn, this did not happen frequently, but it did on occasion when an interesting flow of conversion was happening and I believe I jumped in out of excitement. I suspect this because I understand I do this occasionally in everyday conversions and I am always trying to aware of it. I also noticed I tried to hand back the conversion to the interviewee when I was aware of my interruptions. These interruptions may have had the impact of the interviewee losing their train of thought, or just not sharing as much information as they intended to. The second thing I noticed was that I did not always ask follow-up questions or probe when I had the opportunity to do so. I know some of the times I did not follow up with another question when I sensed the interviewee was finished, but other times may have been due to my novice status as an social science researcher. While I feel that these limitations may have caused me to miss some information, I am confident that the information the interviewees did share with me was useful and reliable.

Integration. I was planning on cross-referencing the interviewees' explanations on the variation and natural selection with their CINS responses to find similarities and differences. The interviewees did talk about some of same concepts found on the CINS, however the open ended nature of the interview questions often resulted in complex

answers which did not translate to the CINS questions. Therefore, I was only able to compare what were the most difficult concepts for both the participants and the interviewees.

Limitations outside this study involved both theoretical and empirical issues. One theoretical limitation concerned the lack of a definition for narrative effect, Norris et. al, 2005 described narrative effect as an increase in student understanding, memory and interest, but at another point in the paper, narrative effect appears to be any advantage of narratives, such as increased reading speed, and absorption, and increased plausibility and persuasiveness of information. These later advantages may be connected to increased understanding, memory and interest, but Norris et al., 2005 does not discuss them in this manner. Therefore it appears that the authors feel that any advantage is part of a narrative effect. Klassen mentions narrative effect, and refers to Norris et al, 2005 but does not supply his own definition (2009). One empirical issue was with effect size. This was a problem during the interpretation of question QR1. As discussed is Chapter Four, effect size should be reported in an effort to determine the practical significance of research. In spite of this, most research does not report effect size or if they do, often do not explain what method of calculation was used. It is difficult to evaluate this study with previous research when there are no common metrics for comparison. The implications of these limitations are considered in the next section.

Implications

The implications discussed in this section emerged from the research results and from the limitations. What materialized were thoughts about how to improve the MP lesson plan, and considerations for future research.

MP lesson. The MP lesson was designed to explicitly expose the students to past scientific ideas that are now known misconceptions. As mentioned in the results section, the two most frequently held misconceptions that were explicitly discussed had statistically significant declines in the treatment group. It would be helpful if the other most frequently held misconceptions were also explicitly discussed in the lesson. This would include misconceptions involving variation, differential survival and population change which include incorrect ideas about species and competition at the population level. In addition, participants had a great deal of trouble with the idea of random mutations causing variation. This concept should be explicitly covered in this lesson or before. The distinction is important to make because the DeVries theory of mutation is explored in the MP lesson and some participants had trouble separating the two ideas.

It would be good to put a hands-on component back into the lesson. The instructors' survey revealed they were not sure the BugHunt simulation was needed and it was not used. However results show that the participants had difficulty with the concept of differential survival and a large portion of both the control and treatment group requested hands-on activities. There are other simulations involving the peppered moths such as "A Birds Eye View of Natural Selection" at peppermoths.weebly.com that could be used. This simulation also has the advantage of talking about Kettlewell's

research. I suspect that one of the issues for the instructors with incorporating a hands-on activity was time, they mentioned about how much information was packed into the three days. There would have to be a reorganizing of the material presented during class to make room for the hands-on activity. A flipped classroom should be considered so all of the material can be covered. The students could cover revised PowerPoints that incorporate the story script at home, and then work in groups with the worksheets and discussions during class. This would free up the time needed for both the discussion on species and population change, and for the hands on activity. Another addition to this lesson could be examples of the different peppered moth variations. Some participants felt that specimens would be better than just the slides and they also considered specimens as hands on. The moth specimens would be a natural fit because the introduction to the MP lesson begins by talking about moth collectors and their role in science. In addition to changes to the MP lesson, there are assessment changes that could yield more complete information on the participants' explanations.

If I were to conduct this study again, one change would be to include a set of written open-ended questions in addition to the CINS instrument. These questions would be related to the interview questions on variation and natural selection. This would allow a comparison to be made between the written and oral questions. Another form of assessment would be to have the participants write their own story of the peppered moth phenomenon. Research suggests that narrative writing promotes understanding and retention of scientific concepts (Ritchie, Rigano, & Duane, 2008), and helps students be active participants in their learning. In addition narrative writing can be used as a formative assessment that gives the instructor better insight into what the students are

thinking and learning (Miele, 2010). All these suggestions involve the MP lesson, and they are fine as far as they go, however the limitations section discussed the gaps in this study and the next section discusses possible ways to close these gaps.

Future research. Going forward, future research should consider what has been accomplished in this and other studies on the use of stories, and explore the revealed gaps. The suggestions below look at the some of these gaps and suggest a way forward. This study focused on evaluating learning outcomes and the interviewees' experience with the MP lesson. The results revealed what attributes the interviewees thought were part of the MP lesson and in mysteries in general. These attributes have benefits for student behavior and cognition. Future research should explore the presence and effect of these attributes. For example, the effect of mystery stories on student attention, focus and retention could be considered.

This study was also concerned with the inclusion of narrative elements in both the control and treatment groups. The control group used a historical account, and this account contained some narrative elements. The treatment group contained all the narrative elements. In future research the control group should convey the peppered moth phenomenon in an account that contains a minimum amount of narrative elements. It would be difficult to include a phenomenon in a historical science without some narrative elements. For example, past time would probably be present. The treatment group would still contain all ten narrative elements. This design would help establish which narrative elements are necessary to induce a narrative effect.

The research results show that the interviewees had positive views on mysteries and felt there were many advantages to using a mystery as a teaching tool. This study did not explicitly connect the idea of a mystery to a story. Mysteries could just refer to a question or puzzle and may not necessarily have a story format. Future research should consider a lesson like the MP and compare it to one that only answers a question and is not in the context of a story. The interviewees talked about the advantages of mysteries, which as stated earlier, parallel those of stories. It would be a step forward if there were a connection between mysteries and stories. Another interesting point the interviewees made was that they considered mysteries to be inquiry-based, and this was a positive point for them. Possible future research could compare a mystery based story against a story without a mystery. This would help pinpoint the advantages of mysteries vs. stories.

Final thoughts. This study considered the impact of a historical story on participants' learning outcomes, and it looked at the experiences of the interviewees with the MP lesson. The CINS results show that the use of historical story had a statistically significant positive impact on learning outcomes, including gains in CINS scores and decreases in the use of commonly held misconceptions. The interview responses revealed positive attitudes with the MP lesson and for mysteries in general. For example, the interviewees, as students and future teachers, revealed that a main strength of mystery-based lessons is the inquiry component, which is an important part of science education. The study's results provide a connection between previous work on stories, both theoretical and empirical, and future work. For example, this study through the use of two versions of the MP lesson, has demonstrated that there is a systematic method to empirically test the efficacy of stories which can be employed in future research.

APPENDIX A

HSIRB Approval Forms

WESTERN MICHIGAN UNIVERSITY

Human Subjects Institutional Review Board

Date: October 21, 2013

To: David Rudge, Principal Investigator
Janice Fulford, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair Amy Naugle

Re: HSIRB Project Number 13-10-18

This letter will serve as confirmation that your research project titled "Assessing the Impact of a Narrative-based Historical Case Study on Student Learning of Natural Selection" has been **approved** under the **expedited** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., ***you must request a post approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study***)." Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: October 21, 2014

251 W. Walwood Hall, Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276

Informed Consent Form
Western Michigan University
Mallinson Institute for Science Education

Principal Investigator: Dr. David Rudge
Student Investigator: Janice Fulford

Title of Study: Assessing the Impact of a Narrative-based Historical Case Study on Student Learning of Natural Selection

You have been invited to participate in a research project titled " Assessing the Impact of a Narrative -Based Historical Case Study on Student Learning of Natural Selection" This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

What are we trying to find out in this study?
The purpose of this study is to determine the impact of a historical case study on student learning of natural selection.

Who can participate in this study?
Any student enrolled in BIOS 1700 can participate in this study.

Where will this study take place?
This study will be conducted on the campus of Western Michigan University. Assessments will be completed in class. Interviews will be conducted in Wood Hall.

What is the time commitment for participating in this study?
The time commitment for participating in this study consists of two 20 minute sessions during scheduled class time for completing assessments. Additionally, you may be asked to participate in a 30 minute interview outside of class. All participation will be complete within one month of the start of the study.

What will you be asked to do if you choose to participate in this study?
If you choose to participate in the study you will be asked to allow the data collected during the pre- and post-assessment, conducted during class, to be used for research purposes. You may also be asked to participate in an interview regarding your impressions of the Mystery Phenomenon Unit.

What information is being measured during the study?
Your responses to the assessment instruments will be collected and coded. The interviews will be recorded using a digital recorder and transcribed. The responses to the instrument and the interviews will then be analyzed to determine the impact of the lesson.

What are the risks of participating in this study and how will these risks be minimized?
The risks involved in participating in this study are minimal. In order to mitigate any potential risk; no identifiable information will be shared with anyone but the primary researchers, with the exception of interview quotes. Also, you will not be asked to talk about any information that you are uncomfortable with sharing. PARTICPATION IN THIS STUDY WILL IN NO WAY IMPACT YOUR GRADE FOR THIS COURSE.

What are the benefits of participating in this study?
There are no direct benefits to you as a participant. The main benefit for the field of science education is that the study may add to the available pool of instructional approaches that positively impact student understanding of the important biological concept of natural selection.

Are there any costs associated with participating in this study?
There are no costs associated with this project outside of traveling costs associated with coming to the campus of Western Michigan University.

Is there any compensation for participating in this study?
There will be extra credit given to subjects who participate in the interview.

Who will have access to the information collected during this study?
Only the researchers will have access to the information collected in this study during data analysis. No identifiable information concerning the participants will be shared; only aggregate data from the CINS and redacted quotes from the interviews will be included in all forms of dissemination.

What if you want to stop participating in this study?
You can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either academically or personally if you choose to withdraw from this study.

The investigator can also decide to stop your participation in the study without your consent.

Should you have any questions prior to or during the study, you can contact the student investigator, Janice Fulford at 269-447-6781 or janice.m.fulford@wmich.edu.

WESTERN MICHIGAN UNIVERSITY
H. S. I. R. B.
Approved for use for one year from this date:

OCT 2 1 2013

x Amy Naugle
HSIRB Chair

You may also contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

____

Please Print Your Name

______________________________ ______________________________

Participant's signature Date

___ Check here if you are willing to participate in 30 minute interview. Please provide an email that you can best be reached at for scheduling the interview in the space below.

Email: ______________________________

Western Michigan University
H.S.I.R.B.
Approved for use for one year from this date:

OCT 21 2015

HSIRB Office

Informed Consent Form
Western Michigan University
Mallinson Institute for Science Education

Principal Investigator: Dr. David Rudge
Student Investigator: Janice Fulford

Title of Study: Assessing the Impact of a Narrative-based Historical Case Study on Student Learning of Natural Selection

You have been invited to participate in a research project titled " Assessing the Impact of a Narrative -Based Historical Case Study on Student Learning of Natural Selection" This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

What are we trying to find out in this study?
The purpose of this study is to determine the impact of a historical case study on student learning of natural selection.

Who can participate in this study?
Any student enrolled in BIOS 1700 can participate in this study.

Where will this study take place?
This study will be conducted on the campus of Western Michigan University. Assessments will be completed in class. Interviews will be conducted in Wood Hall.

What is the time commitment for participating in this study?
The time commitment for participating in this study consists of two 20 minute sessions during scheduled class time for completing assessments. Additionally, you may be asked to participate in a 30 minute interview outside of class. All participation will be complete within one month of the start of the study.

What will you be asked to do if you choose to participate in this study?
If you choose to participate in the study you will be asked to allow the data collected during the pre- and post-assessment, conducted during class, to be used for research purposes. You may also be asked to participate in an interview regarding your impressions of the Mystery Phenomenon Unit.

Western Michigan University
H.S.I.R.B.
Approved for use for one year from this date:

OCT 21 2015

HSIRB Office

What information is being measured during the study?
Your responses to the assessment instruments will be collected and coded. The interviews will be recorded using a digital recorder and transcribed. The responses to the instrument and the interviews will then be analyzed to determine the impact of the lesson.

What are the risks of participating in this study and how will these risks be minimized?
The risks involved in participating in this study are minimal. In order to mitigate any potential risk; no identifiable information will be shared with anyone but the primary researchers, with the exception of interview quotes. Also, you will not be asked to talk about any information that you are uncomfortable with sharing. PARTICPATION IN THIS STUDY WILL IN NO WAY IMPACT YOUR GRADE FOR THIS COURSE.

What are the benefits of participating in this study?
There are no direct benefits to you as a participant. The main benefit for the field of science education is that the study may add to the available pool of instructional approaches that positively impact student understanding of the important biological concept of natural selection.

Are there any costs associated with participating in this study?
There are no costs associated with this project outside of traveling costs associated with coming to the campus of Western Michigan University.

Is there any compensation for participating in this study?
There will be extra credit given to subjects who participate in the interview.

Who will have access to the information collected during this study?
Only the researchers will have access to the information collected in this study during data analysis. No identifiable information concerning the participants will be shared; only aggregate data from the CINS and redacted quotes from the interviews will be included in all forms of dissemination.

What if you want to stop participating in this study?
You can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either academically or personally if you choose to withdraw from this study.

The investigator can also decide to stop your participation in the study without your consent.

Should you have any questions prior to or during the study, you can contact the student investigator, Janice Fulford at 269-447-6781 or janice.m.fulford@wmich.edu. You may also

Western Michigan University
H.S.I.R.B.
Approved for use for one year from this date:

OCT 21 2015

HSIRB Office

contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study. This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

____

Please Print Your Name

__________________________ ________________________

Participant's signature Date

___ Check here if you are willing to participate in 30 minute interview. Please provide an email that you can best be reached at for scheduling the interview in the space below.

Email:__________________________

APPENDIX B

Recruitment Scripts

Class Invitation Script: ☺

I want to thank ____________________(instructor's name) for letting me have a few moments of your time. I am conducting a research study as part of my graduate program at Mallinson Institute for Science Education here at Western Michigan University. I am studying how a historical case-study impacts learning. This class is starting a unit called the Mystery Phenomenon and as part of your normal course work you will be taking a pre- and post -assessment. If you choose to participate in the study the data collected from the pre- and post-assessments will be used for research purpose. In addition to the data from the assessments I will be conducting interviews about this unit. If you do participate in the interview, you will receive EC. However, choosing not to participate in any portion of the research will not have a negative impact on your course grade. Your instructor will not know if you choose to participate. Please take some time to review the consent form, and then I will answer any questions you may have.

Thank-you for your time ☺

Interview Email Fall 2013

Dear ___________

Thank you for your interest in my research project examining student understanding of natural selection. You indicated that you would be willing to participate in a face to face interview on your informed consent form. This interview will last about 30 minutes. These interviews will be conducted here on campus in Wood Hall. You are not required to participate in the interview. However, you may help to increase what is known about the impact of narrative-based historical case studies on student understanding of natural selection, which an important biological concept. This information could be valuable for improving science education.

If you are still interested in participating please again carefully review the consent form (attached to this email). It is important that you review this form carefully as it contains critical information regarding the study. After reviewing the consent form, if you have any questions please feel free to contact me with any questions and concerns by phone at 269-447-6781 or janice.m.fulford@wmich.edu

If you would still like to participate in the interview please send me two dates/times that work best for your schedule. I will send you a confirmation email with a room and time for the interview. Thank you.

Sincerely,

Janice Fulford

Doctoral Student

Mallison Institute for Science Education

Western Michigan University

janice.m.fulford@wmich.edu

269-447-6781

APPENDIX C

Mystery Phenomenon Lesson Scripts

MP Lesson Original Script 1

Slide 1

- As was just pointed out, <u>every</u> population in nature is composed of organisms that differ from one another.
- A central problem at the turn of the twentieth century was that of accounting for both the origin and the incidence of variation in nature, both within groups and between groups.
- We will be discussing variation in the context of evolutionary change. One key distinction for us throughout this unit will be the distinction between *macro-evolution* and *micro- evolution*.
- Can anyone explain to us the difference between, macro- and micro- evolution?

Slide 2

- **Evolution** in general refers to changes in species over time.
- **Macro-evolution** refers to large scale irreversible changes that are thought to take place over large spans of time (e.g. thousands of generations and millions of years), such as speciation (the origin of new species) and extinction (the loss of species). These changes occur in the biological hierarchy at the level of species and larger taxonomic groups.
- **Micro-evolution**, in contrast, refers to reversible evolutionary changes that occur in relatively shorter periods of time (e.g. tens of generations, over hundreds or thousands of years), such as changes in the composition of a population over time. Thus whereas macroevolution is evolutionary change at the level of <u>species</u> and larger taxonomic groups; microevolution occurs at the level of <u>populations</u>. Because species are composed of populations a central question we will want to consider as we go through this unit is what, if anything, is the relation between micro- and macro- evolution.
- I now want to introduce you to our Mystery Phenomenon and I'd like you to consider whether it is or is not an example of microevolutionary change.

Slide 3

- Here is a photograph of the typical form of a moth known as *Biston betularia*. It is a very common moth throughout Continental Europe and a member of a night flying order of the Lepidoptera known as the Heterocera. It's pale appearance actually represents a very complex pattern, and as you can see,

Slide 4

- when it rests on the surface of lichen covered tree trunks in rural areas during the day it is very difficult to detect.
- Let me now introduce our mystery about this moth.

Slide 5

- The moth was first characterized in 1766 by the British naturalist, Moses Harris.

Slide 6

- Up until the early nineteenth century, this moth was believed to have one pale form, known as **f. *typica***.

Slide 7

- In 1848 a different dark colored form was discovered by an amateur lepidopterist, which was named *carbonaria*.

Slide 8

- Here's a photograph of this form. [8]

Slides 9-19

- Moth collectors, like stamp and coin collectors today, were very excited by this finding and continued to look for more specimens. Interestingly, over time they discovered more and more examples of it.

Slide 20

- This is when biologists started to get interested in it, because people noticed that only a hundred years time not only was this dark form being found in an increasing larger area, but whole populations were shifting from being composed primarily of the light form to primarily of the dark form. [20]

Slide 21

- Here's an image of a frequency map of the two forms of the moth. The circles are pie graphs that represent the relative frequencies of the pale and dark forms. When the circle is completely white, it means the entire population is composed of the pale form; when it is half white, only half of the population in that locality is white, the rest are dark. Does any one notice a pattern? (*The rise in frequency seems to be limited to a particular part of the country*).
- Scientists were fascinated not only by the fact that the dark form was becoming more common, but also that this increase only seemed to be occurring in certain areas throughout Britain and also Continental Europe. What do they have in common? Does anyone recall the major change that was taking place in the British Isles and Continental Europe from about 1820 onwards? (*The Industrial Revolution*.)

Slide 22

- For the most part the areas where the dark form is becoming more common are all downwind of industrial sites. Unlike the pale lichen covered trees in rural areas, the forests in these areas are blanketed with soot that kills off the lichen cover.

Slide 22

- Here you see a depiction of what happened to the countryside outside of Manchester. So there appears to be a correlation between air pollution and an increase in the frequency of these dark moths. Moreover, naturalists noticed a similar change was occurring in at least fifty other species.

Slide 23

- Now let me briefly share with you the results of some initial studies of the genetics of the particular moth we have been discussing, *Biston betularia*. Breeding experiments conducted in the early 1940s established that there was only one gene responsible for the dark coloration, and moreover, that it not only made moths darker but they also appeared to be "hardier" or physiologically stronger than the pale form.

Slide 24

- Let's go back to our original question. In view of what I've told you so far, does this appear to be an example of micro or macro-evolutionary change? *Because this is a population level change and has a genetic basis, it appears that it might be an example of micro-evolution.*
- Our mystery phenomenon, is the question of why industrialization per se would cause this change.

Slide 25

- What I'd like you to do now is split up into groups and develop a theory about why you think this change is occurring.

MP Lesson Original Script 2

Slide 1

- Begin by mentioning that the balance of today will be devoted to a continuation of our discussion about the mystery phenomenon and how we might account for it.

Slide 2

- Remind them that the mystery phenomenon concerns a moth species (*Biston betularia*) that until early in the nineteenth century was thought to only have one form, f. typical.

Slide 3

- In 1848 a dark form of this moth was discovered outside of Manchester

Slide 4

- and that by 1950 the dark form had not only spread to other areas,

Slide 5

- but was also becoming extremely common in some of these areas -as much as 90% in the space of only a hundred years as indicated in pie graphs of frequencies.

Slide 6

- which is an extremely rapid change from the standpoint of evolution. Remind them that we discovered that there seems to be a correlation between areas where the dark form has become more common and pollution centers.

Slide 7

- Point out that if you visited unpolluted areas, trees in these areas have a pale appearance owing to the presence of lichens that cover them,

Slide 8

- whereas in polluted areas this lichen cover has been completely killed off by the pollution and the trees are darkened by soot.

Slide 9

- Draw their attention to how some of us suggested that the change might have to do with the obvious color difference between the two moths,

Slide 10

- and in particular how difficult it is to locate the pale moth in rural forests.

Slide 11

- and the dark form in polluted areas,

Slides 12,13

- Document this with photographs in these two settings and invite students to see if they can spot both moths in each photograph. Point out also that some of us suggested that the alleged physiological differences between the two forms of the moth might also play a role - perhaps the dark form is more resistant to toxins in the pollutants.

Slide 14

- So the mystery phenomenon is accounting for why the dark form is becoming more common in the polluted forests.

Slide 15

- With this summary in place, tell students that we will explore the tentative answers they suggested in more depth, and in particular, three different theories that might account for this change.
- Ask students to remind the class what these three theories were: (*1). Physiological Adaptation to Changed Circumstances*, (2) *Natural Selection* and (3) *Mutation*. [

Slide 16

- Point out that each of these three theories was once seriously considered by scientists, and that you will now briefly review these from the perspectives of the scientists who proposed them: (*1). Lamarck's Theory of Acquired Inheritance*, (2) *Darwin's Theory of Natural Selection* and (3) *De Vries' Mutation Theory*.

Slide 17

- We will discuss both the theory in general and consider how subsequent workers proposed it applies to the "Mystery Phenomenon".
- <u>Lamarck's Theory of Acquired Inheritance</u>. Share the basic tenets of Lamarck's theory using the intuitive example of how giraffe's necks got longer.
- Emphasize that no one doubts that bodies can adapt within limits to their surroundings (e.g. plants grow towards sources of light, people tan on the beach, people who work out in the rec. center regularly can get more muscular). What is controversial about this theory is whether these sorts of changes that take place in the life time of an organism can be passed on to its offspring.

Slide 18

- Point out that a scientist (Nicholas Cooke) suggested just like them that perhaps the dark form was becoming more common due to a physiological adaptation on the part of individual organisms to their environment- and in the case of the peppered moth, a general darkening caused by pollution and increased humidity has led the moths to adapt to their environment by becoming darker in order to hide from birds.

Slide 19

- Darwin's Theory of Natural Selection. Share the basic tenets of Darwin's theory using the intuitive example of antelope becoming quicker as a result of predation by cheetahs against slow antelope.
- Emphasize the contrasting claims made by Lamarck and Darwin - whereas Lamarck claimed these changes take place within the life of an organism, Darwin argued that populations have variation to start with and that the change that occurs is at the population level - in terms of shifting frequencies of the various forms and not changes at the individual level.

Slide 20

- Point out that a amateur lepidopterist (James Tutt) suggested just like them that perhaps the dark form was becoming more common due to the selective advantage of being dark in polluted environments.

Slide 21

- Draw attention to how the ecological geneticist E.B. Ford built off of Tutt's theory by combining his insights about the physiological advantage of the dark form by suggesting that the spread was due to the physiological difference and the limit to the spread was imposed by the obvious handicap of being dark colored in an unpolluted environment.

Slide 22

- De Vries' Mutation Theory. Share the basic tenets of DeVries mutation theory using results of his famous experiment on the evening primrose.
- Emphasize that De Vries disputed that species level differences could occur with reference to the accumulation of slight changes as Darwin proposed, but instead were the result of dramatic changes due to mutation.

Slide 23

- Point out that the scientist J. W. Heslop-Harrison just like some of them believed this might account for the increase in the dark form, and that he was convinced that lead salts in the soot might have mutagenic properties.

Slide 24

- Cooke's Theory. Emphasize that Cooke's theory suggests the change was gradual, and indirectly caused by pollution. On Cooke's theory, the darkening of the environment caused each individual moth in the polluted environment to adapt and these changes have been inherited by their offspring.

Slide 25

- Tutt and Ford's Theory. Tutt and Ford's theory also suggests the change was gradual and an indirect result of the pollution. However, they contended the dark form arose as a result of mutation *completely by chance* not because of either the direct or indirect effect of pollutants in the environment. On their view, the dark form has *always* been present at a low frequency in the population as a result of

recurrent mutations. The pollution did <u>not</u> cause the mutation to occur, it simply changed the environment in a way that now favors the survival of the dark form relative to the pale form and as a consequence, its numbers are rising. On this view, it is <u>not</u> the individual that is changing– it is the composition of the population over time.

Slide 26

- <u>Heslop-Harrison's Theory</u>. Heslop-Harrison's theory suggests the change was very dramatic. Iron salts in the pollution *directly* caused individual moths to mutate and this mutation has been passed down to their offspring.

Slide 27

- *What is a theory?* Discussion of a theory

Slide 28

- *How do scientists choose between competing theories?* Comparison of theories

Slide 29

- Kettlewell investigation introduction
- Point out to students that H.B.D. Kettlewell attempted to experimentally determine whether birds in fact prey upon the moths in nature. Point out that his research involved releasing marked moths of both types in a polluted and an unpolluted setting.

Slide 30

- He then attempted to recapture as many as possible, the presumption being that differences in recapture rate would reflect differences in their ability to avoid predators during the time between release and recapture. The recapture techniques he used involved both a mercury vapour light trap and assembling traps. Point out also that he attempted to film bird predation as well.

Slide 31

- <u>The Kettlewell movie "Evolution in progress</u>"
- Show the movie to students and answer any questions they might have.

MP Lesson Modified Script 1

Transition to the PowerPoint

- We just finished an activity that pointed out the variation in traits in our classroom, and we are not unique. All populations have individuals that are different from each other and we are going to continue our exploration of variations by looking at how change may happen.

Slide 1

- An important scientific problem at the turn of the 20^{th} century was understanding the origin and occurrence of variation in and between groups.
- We are going to be looking at variation in terms of evolutionary change, so we need to understand the difference between *macro-evolution* and *micro-evolution.*
  Evolution refers to changes in species over time
  Marco-evolution refers to large-scale irreversible changes that are thought to take place over large spans of time (Thousands of generations and millions of years), such as speciation (origin of new species) and extinction (loss of species). These changes occur at the species level and larger taxonomic groups.
  Micro-evolution refers to reversible evolutionary change that occur in shorter time spans (tens of generations; over hundreds or thousands of years), such as changes in a population's composition. These changes happen at the population level

- As we go through this unit about the mystery phenomenon, I'd like you to consider if it is an example of micro-evolutionary change.

Slide 2

- People have been fascinated with the variations in moths and butterflies for many centuries. Collections such as this one show variations in species.

Slide 3

- In Britain, people are avid collectors; they are interested in the variety of butterflies and moths, they enjoy being outdoors, it is a social activity that the young and old alike can enjoy, and they are very active citizen scientists. They collect data that is used for scientific and conservation purposes.

Slide 4

- Butterfly and moth collecting in Britain has been going on for centuries, and for much of the same reasons as today, it was a very social, and at times competitive

activity. People often started collecting as children continued to collect their entire lives.

- Collectors traded specimens, shared information and kept amazing records. Some collectors were only interested in having the most complete collections with all the odd and rare specimens. Others were interested in understanding the natural history of species and they contributed a great deal to the knowledge of butterflies and moths.

Slide 5

- Our mystery phenomenon is about one particular British moth, *Biston belularia*, a night-flying moth, which was described by collector and naturalist, Moses Harris in 1766. It was thought that there was just one form of the moth, ***f. typica***.

Slide 6

- Notice how *typica* looks when resting on this lichen-covered tree trunk.

Slide 7

- We can see that around the mid-nineteenth century *typica* had a fairly wide distribution around Britain.

Slide 8

- In 1848 an amateur entomologist found a dark form of the moth, which was named ***f. carbonaria.***

Slide 9

- One collector who knew about and was interested in *carbonaria* was J.W. Tutt.
- Tutt was a high energy, amateur entomologist who was an avid collector from his youth. He could be seen in the field with his pockets stuffed with tins full of butterflies and moths; He believed a species could not be understood from only one or two specimens.
- He was a school-master who loved learning and teaching but he had an abrupt personality and he often offended people with his lack tolerance for any opinion he thought was wrong. He however was intellectually honest, and was quick to admit when he was wrong.
- Tutt was also a very prolific writer, sharing information as founder and editor of a journal for amateur entomologists: one that is still in publication today. His book, ***British Moths*** instructs young collectors to do more than form collections; they should study how species change and form variations. Tutt also wrote about the increase in *carbonaria*.

- Tutt worked endlessly on his observations, collections and writings, looking for answers to so many great questions. His relentless pace may have contributed to his early death in 1911. The question of *carbonaria* was one question he did not get answered.
- Collectors such as Tutt, were excited about *carbonaria,* and more people looked for it.

Slides 10-19

- Over time more specimens were found, and scientists were beginning to show interest in this moth, because in only 100 years the dark form was increasing in frequency and territory.

Slide 20

- Here is a frequency map of the two forms of the moth. A completely white circle represents a population of entirely light moths; a ½ white circle, only ½ the population is light. What pattern can you see about the dark form of the moth? *(The rise in dark form frequency limited to certain part of country).*
- Scientists were fascinated that the dark form was becoming more common, and that it seemed to be occurring only in certain areas of Britain and Continental Europe. What could these areas have in common? Does anyone remember what major change was happening from the early 1800s onward in Britain and Continental Europe? *(The industrial revolution)*
- There appears to be a correlation between air pollution and the increase in frequency of the dark form. Generally, the dark form was becoming more common in areas downwind of industrial areas, such as Manchester.

Slide 21

- Here is a painting of the Manchester countryside before the industrial revolution.
- Notice the dramatic change in just 30 years.

Slide 22

- The potential relationship between industrialization and *carbonaria* was not lost on Tutt. He noted how factories threw out soot and noxious vapors and speculated how these things could have contributed to the increase of *carbonaria*.

Slide 23

- The industrial revolution had a profound effect on the British landscape, not only were changes noticed in *Biston betularia*, naturalists noticed changes in at least fifty other species.

Slide 24

  - So our mystery is this: "Why is the dark form becoming more common in areas downwind of manufacturing center?"

MP Lesson Modified Script 2

Transition to PowerPoint # 2

- We just finished the discussion on the Darwin movie. Darwin proposed how change and variation happens, and that's what you did when you proposed theories about the mystery phenomenon.
- We will continue looking into this mystery.

Slide 1

- So let's review where we left off. *(some students may have missed the first class)*

Slide 2

- *Biston betularia* is the British moth that is the central character in our mystery.

Slide 3

- It was thought that there was only one form of this moth. Until…

Slide 4

- A dark form, *carbonaria* was discovered in 1848 outside of Manchester.

Slide 5

- J.W. Tutt was just one of the collectors and amateur entomologists that was interested in *carbonaria*, he believed the increasing pollution in Britain had a role to play in the increased numbers of the dark form.

Slide 6

- Just as Tutt speculated, there seemed to be a correlation between pollution centers and areas where the dark form was common.
- By 1950 the dark form had spread beyond Manchester and was extremely common in certain areas, up to 90% in some cases, all in a timespan of a hundred years.

Slide 7

- So let's continue exploring the possible causes of this mystery.

Slide 8

- Look at the unpolluted forest, notice how pale it looks; this is because the trees are covered with lichens. This is not the case with the polluted forest where the lichens are dead and the trees are covered with soot.

Slide 9

- Some of us suggested that change in the amount of dark moths might have something to do with the color difference between the two moth forms.
- See how difficult it is to see the light form in the unpolluted forest? Can you see the light form?

Slide 10

- Now look at how difficult it is to see the dark form in the polluted forest. Can you find the dark form?
- Some of us suggested that there might have been physiological differences between the two moth forms that played a role, maybe the dark form is more resistant to toxins in the pollutants. So the mystery continues…Whyis the dark form more common in polluted forests.

Slide 11

- So we are going to look at the theories you suggested during our last class. Can anyone us what those theories are? (1) *Physiological adaptation to Changed Circumstances*,(2)*Natural Selection* and (3) *Mutation.*

Slide 12

- Each of these three theories was once considered by scientists; and we are going to look at each of these theories from the perspective of the scientists who proposed them. We are going to look a the theories in general and how subsequent people applied them to the mystery phenomenon.

- These theories are (1) Lamarck's Theory of Acquired Inheritance, (2) Darwin's Theory of Natural Selection and (3) De Vries' Mutation Theory.

Slide 13

- <u>Lamarck's Theory of Acquired Inheritance.</u> Share the basic tenets of Lamarck's theory using the intuitive example of the giraffe's necks. Emphasize that no one doubts tha bodies can adapt within limits to their surroundings (Plants toward

light, people tan in the sun, or build up muscles). What is controversial about the theory is whether changes in a organism's lifespan can be passed on to offspring.

Slide 14

- In the late 1800s, Nicholas Cooke, suggested individual moths were adapting to the darken environment that was caused by pollution and increased humidity. The individual moth adapted to their environment by becoming darker in order to hide from birds.

Slide 15

- Darwin's Theory of Natural Selection. Share the basic tenets of Darwin's theory using the intuitive example of antelope becoming quicker as a result of predation by cheetahs against slow antelope.
- Emphasize the contrasting claims made by Lamarck and Darwin. Lamarck claimed changes take place within the life of an organism, Darwin argued that populations have variation to start with and that the change that occurs is at the population level, in terms of shifting frequencies of the various forms and not changes at the individual level.

Slide 16

- Remember that J.W. Tutt suggested that the pollution had something to do with the increase of the dark form?
- He thought the dark forms had a selective advantage in the polluted forests due to crypsis. The dark form was better camouflaged.

Slide 17

- E.B. Ford collected butterflies and moths with his father as a child, and became very interested in different varieties. He decided to study biology and genetics as a result of this interest. He was an ecological geneticist at Oxford.
- Ford, like Tutt a ½ century before, passed his interest and knowledge of butterflies and moths along in two books written for collectors. These books remained best sellers for decades.
- He built off of Tutt's theory by suggesting the spread of the dark form was due to a physiological advantage, but the limits of the spread was due to the disadvantage of being dark in unpolluted environments.

Slide 18

- DrVries' Mutation Theory Share the basic tenets of DeVries mutation theory using the results of his famous experiment on the evening primrose.
- DeVries disputed that species level differences could occur by the accumulation of slight changes as Darwin proposed, but were the result of dramatic changes due to mutations.

Slide 19

- Heslop Harrison had a life-long interest in plants and insects. From a young boy, his surroundings were filled with insect cages, tins and other equipment.
- He was an energetic, highly- praised high-school science teacher, whose students credit him with cultivating their interest in the natural world. He earned his doctorate and moved on to a college position.

- Heslop Harrison's scientific work involved genetics in butterflies and moths. As a scientist, he was a loner and did a great deal of work from home. He had an uncompromising personality and he was always convinced of the correctness of his theories and views.
- He thought as DeVries and felt mutations accounted for the increase in the dark form. He was convinced that the lead salts in the soot might have mutagenic properties.

Slide 20

- To sum up the three theories on the mystery phenomenon…

Slide 21

- Cooke's theory suggests the change was gradual, and indirectly caused by pollution. The darkening of the environment caused each individual moth in the polluted environment to adapt and there changes have been inherited by their offspring.

Slide 22

- Tutt's and Ford's Theory also suggests the change was gradual and an indirect result of the pollution. However, they contended the dark form arose as a result of mutation completely by change not because of either the direct or indirect effect of pollutants in the environment.
- From their view, the dark form has always been present at a low frequency in the population as a result of recurrent mutations. The pollution did not cause the mutation to occur, it simply changed the environment in a way that now favors

the survival of the dark form relative to the light form. As a consequence, the dark form is increasing, the population composition is changing, not individual moths.

Slide 23

- Heslop Harrison's theory suggests the change was very dramatic. Iron salts in the pollution directly caused individual moths to mutate and this mutation has been passed down to their offspring.

Slide 24

- With these examples in mind, we are going to consider what is a theory. (Theory Discussion)

Slide 25

- How do scientists choose between competing theories? (Use this slide as a way to transition from the theory discussion to the students' investigation designs)

MP Lesson Modified Script 3

- Transition: Today we continue our discussion about the mystery phenomenon and how evolutionary biologists study phenomena like this.

Slide 1

- Let's review what the Mystery is…

Slide 2

- This mystery involves *Biston betularia*, a moth species that has a light and dark form. Remember that until 1848, only the light form was known, and that in a 100 years of the dark form's discovery, it was the most common form in certain areas.

Slide 3

- These areas were downwind of industrial sites, and we have been trying why industrialization caused this rapid change.
- Last class we discussed 3 possible explanations. Can someone remind the class what Cooke's theory said? What about Heslop Harrison's theory?

Slide 4

- Can someone describe Tutt's and Ford's theory using these images of moths?

Slide 5

- To summarize; the spread of the dark form is due to its physiological superiority over the light form. And the spread is limited to areas darkened by pollution owing to the handicap or dark coloration in unpolluted environments.

Slide 6

- Mystery Phenomenon Part III

Slide 7

- Last class we looked at Kettlewell's Great Tit experiments, he wanted to see if birds had the same difficulty of finding moths as humans do. Kettlewell felt the results of the experiments were successful.
- But because Kettlewell was interested in testing the second part of Ford's theory, the next step was to conduct experiments in unpolluted and polluted settings.

Slide 8

- Kettlewell released marked dark and light forms in a polluted and an unpolluted setting and then attempted to recapture them.
- He presumed that differences in the recapture rates reflected differences in the moth's ability to avoid predators.
- Slide 9
- Kettlewell film clip

Slide 10

- In the unpolluted forest; twice as many light moths were recaptured as the dark moths.

Slide 11

- In the polluted forest; twice as many dark moths were recaptured as the light moths. In addition, Kettlewell observed a selection coefficient of 30% which led him to revise Ford's original theory. He suggested that the color advantage alone may be responsible for the spread of the dark form. (*This illustrates one way theories are modified over time.)*

Slide 12

- In addition to the mark –release-recapture experiments, Kettlewell attempted to film the birds preying on the moths. Tinbergen's films did capture birds selectively feeding on moths.

Slide 13

- We have been talking about different experiments conducted by various scientists. What are experiments?

  (An experiment is a systematic study of the response of a system to some change. Discuss experiments with class, examples include; Heslop Harrison lead salts, Kettlewell' s experiments, Weismann's cutting tails off rats to test Lamarckian theory)

Slide 14

- So are experiments necessary for progress in science?
  (No, observations, surveys and life histories are all tools Kettlewell used to learn about industrial melanism of Biston Betularia. Experiments are only as good as

the underlying assumptions they are built on, many of which are rooted in natural history).

Slide 15

- Have Kettlewell's experiments convinced you that Ford's theory is correct?

(Most students will feel Kettlewell's work established Ford's theory is correct)

Slide 16

- Kettlewell's work is compelling, it does appear that dark moths have a selective advantage in the polluted forests, and the films are a convincing demonstration of bird predation.
- However…Kettlewell's conclusions were not accepted by all scientists. They felt there were issues with the design of the experiments and that the behavior and ecology of the moth was not well known, so these things together cast doubt on the results.
- Other researchers, independent of Kettlewell were also working on industrial melanism, and this work continues today.

Slide 17

- One of the scientists also working on *Biston betularia* was Michael Majerus. He became interested in moths and other insects early in life. At the age of ten he bought Ford's book *Moths* and this fostered an interest in Ford's work and ecological genetics in general. He was an enthusiastic educator and he wrote several books on insects, one on moths.

Slide 18

- Both Kettlewell and Majerus are considered experts on melanism; they each authored books on the subject.
- Majerus critiqued Kettlewell's work, but he was supportive of bird predation as the main selective factor. He therefore conducted experiments aimed at correcting weaknesses in Kettlewell's experimental designs.
- Majerus conducted a six -year experiment using 4864 moths that addressed flaws in Kettlewell's protocols. He concluded that selective bird predation accounts for the frequencies changes in the light and dark forms. He felt Kettlewell's key assumption was correct.

Slide 19

- Michael Majerus was not able to publish his results because he died unexpectedly in 2009 at the age of 53. A team of scientists who have also spent significant time researching the moth, reviewed his results, supported his conclusions and wrote up his research, which they published in 2012.
- They concluded that Majerus' "data provide the most direct evidence yet to implicate camouflage and bird predation as the overriding explanation for the rise and fall of melanism in moths.

Slide 20

- So back to the dark form of the moth. We left off in the 1950s with its frequency over 90% in some locations. Let's see if this is still the case today.

Slide 21

- Pollution was rather dreadful in many parts of Britain (and other places around the globe), and in 1952 London had a horrible smog incident.
  The was a 4 -day smog event that caused at least 4000 deaths, revised reports put the deaths at 12,000. 10,000 became sick due to the smog.

Slide 22

- This prompted the Clean Air Act of 1956. How do you think the Clean Air Act affected the frequency of the dark moth?
- This graph shows the frequency of the dark form declining in Leeds, Manchester and Caldy and York from 1960 through 2000.

Slide 23

- We have been exploring an evolutionary mystery that started at least 165 years ago. We looked at three theories, multiple experiments and researchers and chose what we feel in the best supported solution.

Slide 24

- But I want you to think about this: Is that the Complete Story? Can you ever tell the entire story?

Slide 25

- I want to share one thing that was not mentioned in the three days we talked about the moth.

It has three forms not just two. Is this a problem? Did the researchers know about it, and take it into account in their work? Would we have made a different decision about the phenomenon if we had known?

Slide 26

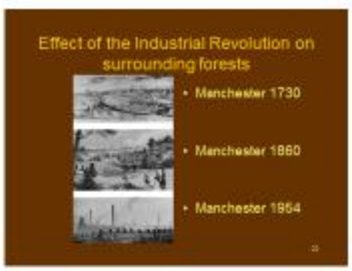
- There were other things not discussed in our account. *Carbonaria* is common in some rural (unpolluted areas), Dark forms of other species have evolved for other reasons and scientific controversies were not covered in detail. In light of this…Slide 27
- What Should Science Teachers Do?
  Have class discuss the four questions on the final slide about simplified versions of scientific accounts.

APPENCIX D

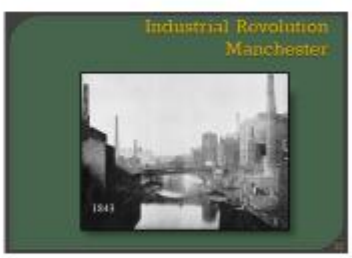
Sample PowerPoint Slide Modifications

Slide 3 Original

Slide 5 Modified



Slide 22 Original
Here you see a depiction of what happened to the countryside outside of Manchester. **So there appears to be a correlation between air pollution and an increase in the frequency of these dark moths.**



Slide 22 Modified
The potential relationship between industrialization and *carbonaria* was not lost on Tutt. **He noted how factories threw out soot and noxious vapors and speculated how these things could have contributed to the increase of *carbonaria*.**

APPENDIX E

Semi-structured Interview Protocols

Fall 2013

Consent questions to be read/answered on tape before interview questions

A. Is it correct that you have signed a consent form which describes the research project and your part in it?
B. Do you understand that your answers to the interview will not impact your course grade in anyway?
C. Do you understand that your personal information will not be released to anyone outside of the researchers working on this project?

---

1. You just finished the 3-day Mystery Phenomenon Unit (MPU), In your opinion, what is the main point of this unit?
    - What about the unit made you think this?
2. The MPU talked about variation. What is your idea of variation?
    - What is an example of variation?
    - How do variations happen?
3. The MPU also talked about natural selection. What in your opinion is natural selection?
    - What is an example of natural selection?
    - How does natural selection work?
4. The name of the unit was the MPU, so what in your opinion was the mystery that you had to solve?
    - Do you feel the class as a whole felt the mystery was solved?
    - Do you personally feel that the mystery was solved?
5. If you think about the entire 3-day MPU, what is your overall impression?
    - What about the unit made you feel this way?
    - What changes would you make to the Mystery Phenomenon?
    - Why do you think this/these change(s) should be made?
6. You have had other mysteries or puzzles to solve in this class, can you tell me about them?
    - How are these earlier mysteries or puzzles the same as the MPU?
    - How are they different?
7. In your opinion, what is the advantage to having these mysteries or puzzles part of the course?

- What is the disadvantage?
- Does the advantage outweigh the disadvantage? Why/why not?

Possible prompts

- Would you expand on that?
- Would you tell me more about that?
- Why do you feel that?

Spring 2014

Consent questions to be read/answered on tape before interview questions

A. Is it correct that you have signed a consent form which describes the research project and your part in it?

B. Do you understand that your answers to the interview will not impact your course grade in anyway?

C. Do you understand that your personal information will not be released to anyone outside of the researchers working on this project?

---

1. You just finished the 3-day Mystery Phenomenon Unit (MPU), In your opinion, what is the main point of this unit?
   - What about the unit made you think this?
2. The MPU talked about variation. What is your idea of variation?
   - What is an example of variation?
   - How do variations happen?
3. The MPU also talked about natural selection. What in your opinion is natural selection?
   - What is an example of natural selection?
   - How does natural selection work?
4. The name of the unit was the MPU, so what in your opinion was the mystery that you had to solve?
   - Do you feel the class as a whole felt the mystery was solved?
   - Do you personally feel that the mystery was solved?
5. If you think about the entire 3-day MPU, what is your overall impression?
   - What about the unit made you feel this way?
   - What changes would you make to the Mystery Phenomenon Unit as a whole?
   - Why do you think this/these change(s) should be made?
6. Could you think specifically about the three PowerPoints?
   - What is your opinion on the topic presented? How did you feel about it?

- What is your opinion about the style of the PowerPoints? The way the topic was presented?
- There were 3 different PowerPoints, how were they tied from one day to the next?
    - What do you think about that?
- What would you change about the presentation? How?

7. You have had other mysteries or puzzles to solve in this class, can you tell me about them?
    - How are these earlier mysteries or puzzles the same as the MPU?
    - How are they different?
8. In your opinion, what is the advantage to having these mysteries or puzzles part of the course?
    - What is the disadvantage?
    - Does the advantage outweigh the disadvantage? Why/why not?

One final question ☺

9. We are going to return to the PowerPoints again, and I want you to think of them all together. Could you perceive of the 3 PowerPoints as a story? Why or why not?
    - What is a story to you?
    - How were the PowerPoints like a story?

Possible prompts

- Would you expand on that?
- Would you tell me more about that?
- Why do you feel that?

REFERENCES

Abd-El-Khalick, F., & Lederman, H.G. (2000). The influence of history of science courses on students' views of nature of science. *Journal of Research in Science Teaching*, *37*(10), 1057-1095.

Abraham, J.K., Meir, E., Perry, J.N., Herron, J. C., Maruca, S., Stal, D. (2009). Addressing undergraduate student misconceptions about natural selection with an interactive simulated lab. *Evolution: Education & Outreach*, *2*(3), 393-404.

Abraham, J. K., Perez, K. E., Downy, N., Herron, J. C., Meir, E. (2012). Short lesson plan associated with increased acceptance of evolutionary theory and potential change in three alternate conceptions of macroevolution in undergraduate students. *CBE Life Science Education, 11,* 152-164.

Allchin, D.K. (2003). Scientific myth-conceptions. *Science Education, 87*(3), 329-351.

Alters, B.J., & Nelson, C.E. (2002). Perspective: teaching evolution in higher education. *Evolution, 56*(10), 1891-1901. Doi:101554/0014-820(2002)056[1891:PTEIHE]2.0.CO;2

American Association for the Advancement of Science (1993). Project 2061 Benchmarks for Science Literacy. Oxford University Press. New York.

Anderson, D.L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching, 39*(10), 952-978.

Arnold, I (2011). John Hattie: visible learning: a synthesis of over 800 meta-analyses relating to achievement. *International Review of Education, 57,* 219-221.

Arons, A.B. (1989). Historical and philosophical perspectives attainable in introductory physics courses. *Educational Philosophy and Theory, 20,*(2), 13-23.

Baretta, L., Tomitch, L. M.B., MacNair, N., Lim, V.K., Waldie, K.E. (2009). Inference making while reading narrative and expository texts: an ERP study. *Psychology & Neuroscience*, 2(2), 137-145.

Bartlett, F.C. (1932). *Remembering*. Cambridge, England: Cambridge University Press.

Battisti, B.T., Hanegan, N., Sudweeks, R., Cates, R. (2010). Using item response theory to conduct a distracter analysis on conceptual inventory of natural selection. International Journal of Science and Mathematics Education, 8, 845-868.

Beardsley, P.M., Stuhlsatz, M. A., Kruse, R.A., Bloom, M.V., Westbrook, A.L. (2011). Design and evaluation of an evolution and medicine high school curriculum intervention. Paper presented at 2011 Annual International Conference of the National Association for Research in Science Teaching. Orlando, Florida.

Bishop, B.A., & Anderson, C.W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415-427.

Blachowicz, J. (2009). How science textbooks treat scientific method: a philosopher's perspective. *British Journal of Philosophy of Science, 60,* 303-344. doi:10.1093/bjps/axp011

Black, J.B., Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 267-375.

Cherryholmes, C.H. (1992). Pragmatism and scientific realism. *Educational Researcher*, *21*(6), 13-17.

Clary, R. & Wandersee, J. (2008). Marquee fossils. *The Science Teacher, 75*(1), pp. 44-50.

Cleland, C.E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*, *69*(3), 447-451. Retrieved from http://www.jstor.org/stable/10.1086/342455

Clough, P. M. (2011). The story behind the science: bringing science and scientists to life in post-secondary science education. *Science and Education, 20*, 701-717.

Cook, L. M., Grant, B. S., Saccheri, I. J., & Mallet, J. (2012). Selective bird predation on the peppered moth: The last experiment of Michael Majerus. *Biological Letters, 8*(4), 609-612.

Conant, J.B. (1947). *On Understanding Science.* Yale University Press, New Haven.

Conant, J.B. (1951). *On Understanding Science: An Historical Approach*, New American Library, New York.

Conant, J.B. (1957*). Harvard Case History in Experimental Science*, Vols. 1 & 2, Harvard University Press, Cambridge.

Creswell, J. W. (1994). *Research Design: Qualitative & Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2007). *Qualitative Inquiry & Research Design.* 2nd Ed. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2008). *Research Design: Qualitative, Quantitative and Mixed Method Approaches.* 3nd Ed. Thousand Oaks, CA: Sage Publications.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., Hanson, W. E. (2003). In V. Plano Clark & J. Creswell (Eds.), The Mixed Methods Reader. (pp. 161-175). Thousand

Oaks, CA: Sage Publications. (Reprinted from *Handbook of Mixed Methods in Social and Behavioral Research*, by Tashakkori & Teddlie (Eds), (2003).

Demastes, S.S., Settlage, J., & Good, R.G., (1995). Students' conceptions of natural selection and its role in evolution: cases of replication and comparison. *Journal of Research in Science Teaching, 32*(5), 535-550.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher, 35*(3), 125-129.

Driver, R., & Easley, J. (1978). Pupils and paradigms: a review of literature related to concept development in adolescent science students. *Studies in Science Education, 5*, 61-84.

Driver, R., & Oldham, V. (1986). A constructivist approach to curriculum development. *Studies in Science Education, 13*(1), 105-122.

Duschl, R.A. (1985). Science education and philosophy of science: twenty-five years of mutually exclusive development. *School Science and Mathematics, 87*(7), 541-555

Egan, K. (1986). *Teaching as storytelling: An alternative approach to teaching and curriculum in the elementary school.* Chicago: University of Chicago Press.

Evans, E .M. (2008). Conceptual change and evolutionary biology: a developmental analysis. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change*, 263-294. New York, Routledge.

Evans, P. L., & Anderson, D. L. (2013). The conceptual inventory of natural selection a decade later: development and pilot testing of a middle school version leads to revised college/high school version. Paper presented at 2013 Annual International

Conference of the National Association for Research in Science Teaching. Rio Grande, Puerto Rico.

Feilzer, M.Y. (2010). Doing missed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research 4*(1), 6-16. Doi: 10.1177/1558689809349691

Fulford, J.M., Rudge, D.W. (2016). The Portrayal of Industrial Melanism in American College General Biology Textbooks. Science & Education. doi 10.1007/s11191-016-9820-z

Gee, H. (2000. *In Search of Deep Time.* New York, The Free Press.

Gould, J. (2002). The Structure of Evolutionary Theory. Cambridge, MA. The Belknap Press of Harvard University Press.

Gregory, T.R., (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution Education Outreach*, *2,* 156-175.

Hake, R. R. (1998). Interactive engagement vs. traditional methods: a six thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74.

Hattie, J. (2008). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.* London. UK. Routledge.

Heilbron, J. L. (2002). History in science education with cautionary tales about the agreement of measurement and theory. *Science & Education. 11,* 3231-331.

Herreid, C.F. (1998) What Makes a Good Case? *Journal of College Science Teaching* 27(3): 163–165.

Herrnstein Smith, B. (1980). Narrative versions, narrative theories. *Critical Inquiry: On Narrative, 7*(1), 213-236.

Hvolbek, R.H. (1993). Teaching the history of science. *Social Education, 57*(7). 384-387.

Irwin, A.R. (2000). Historical case studies: teaching the nature of science in context. *Science Education, 84*(1), 5-26.

Jeffares, B. (2008). Testing times: regularities in the historical sciences. *Studies in the History and Philosophy of Biology and Biomedical Science, 39*, 469-475.

Jenson, M. S., & Finley, F. N. (1995). Teaching evolution using historical arguments in a conceptual change strategy. *Science Education 79*(2), 147- 166.

Jenson, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching 33*(8), 879-900.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher 33*(14), 14-26. Doi: 10.3102/0013189X033007014

Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education, 17*, 27-47. doi:10.1007/s11191-007-9075-9

Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*, 1313-1340. doi:10.1007/s11191-008-9171-5

Kettlewell, H. B. D. (1961). Evolution in progress. (Cinematographer N. Tinbergen)—sound popularization of the selection experiments film distributed by British Universities Film and Video Council, London.

Kim, S.Y., & Irving, K.E. (2010). History of science as an instructional context: student learning in genetics and nature of science. Science & Education, 19, 187-215.

Kitcher, P. (1985). Darwin's achievement. In N. Rescher (ed) *Reason and Rationality in Natural Science: a group of essays.* Lanham, MD. University Press of America.

Klassen, S. (2009). The construction and analysis of a science story: a proposed methodology. *Science & Education, 18,* 401-423.Kubli, F., (2001). Can the theory of narratives help science teachers be better storytellers. *Science & Education*, *10,* 595-599.

Klassen, S. (2010). The relation of story structure to a model of conceptual change in science learning. *Science & Education*, *19*, 305-317. doi:10.1007/s11191-009-9212-8

Kubli, F. (2001). Can the theory of narratives help science teachers be better storytellers. *Science & Education*, *10*, 595-599.

Kuhn, T.S. (1977). *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: The University of Chicago Press.

Leech, N.L., & Onwuegbuzie, A.J. (2009). A typology of mixed methods research designs. Quality and Quanity. 43, 265-275. doi 10.1007/s11135-007-9105-3

Lennox, J.G. (1993). Darwin was a teleologist. *Biology and Philosophy, 8,* 409-421. Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S. & Busick, M.D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. National Center for Special Education Research. U.S. Department of Education.

Maher, J.M., Markley, J. C., & Ebert-May, D. (2013). The other half of the story: effect sixe analysis in quantitative research. *CBE Life Science Education, 12*(3) 345-351.

Majerus, M.E.N. (1998). *Melanism: Evolution in action*. Oxford: Oxford University Press.

Majerus, M.E.N. (2005). The peppered moth: Decline of a Darwinian disciple. In Fellowes, M.D.E., Holloway, G.J., & Rolff, J. (Eds.), *Insect Evolutionary Ecology* (pp. 371-396). Cambridge, MA: CABI Publishing.

Majerus, M.E.N. (2009). Industrial melanism in the peppered moth, Biston betularia: An excellent teaching example of Darwinian evolution in action. *Evolution Education Outreach, 2,* 63-74.

Mandler, J.M. & Goodman, M.S. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior*, *21*, 507-521.

Mandler, J.M. & Johnson, N.S. (1977). *Cognitive Psychology, 9,* 111-151.

Marshall, C. & Rossman, G. (2011). *Designing qualitative research 5ed.* Thousand Oaks, CA: Sage Publications, Inc.

Matthews, M.R. (1994). *Science Teaching: the role of history and philosophy of science*. Routledge. New York.

Matthews, M.R. (2001). How pendulum studies can promote knowledge of the nature of science. *Journal of Science Education and Technology, 10*(4), 359-367.

Mayr, E. (1961). Cause and Effect in Biology. *Science, New Series 134*(3489) 1501-1506.

McComas, W.F. (2008). Seeking historical examples to illustrate key aspects of the nature of science. *Science & Education, 17*, 249-263.

Metz, D., Klassen, S., McMillan, B., Clough, M., & Olson, J. (2007). Building a foundation for the use of historical narratives. *Science & Education, 16*, 313-334.

Millis, K K., & Graesser, A .C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *The Journal of Memory and Language, 33*, 583-599.

Mintzes, J.J., Wandersee, J.H., & Novak, J.D. (1998). Teaching science for understanding: *A Human Constructivist View*. San Diego, CA. Academic Press.

Monk, M., & Osborne, J. (1997) “Placing the history and philosophy of science on the curriculum: a model for the development of pedagogy” Science & Education 81: 405-424.

Nehm, R.H. & Reilly, L. (2007). Biology majors’ knowledge and misconceptions of natural selection. *BioScience, 57*(3), 263-272.

Nehm, R.H. & Schonfeld, I.S. (2008). Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45*(10), 1131-1160.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*: *Evidence of Common Ancestry and Diversity and Natural Selection.* Achieve, Inc. on behalf of twenty-six states and partners that collaborated on the NGSS.

Norris, S.P., Guilbert, S.M., Smith, M.L., Hakimelahi, S., & Phillips, L.M. (2005). A theoretical framework of narrative explanation in science. *Science Education*, *59*(4), 535-554.

Patton, M. (1990). *Qualitative evaluation and research methods*. Beverly Hills, CA. Sage Publications.

Potts, G., Keenan, J., Golding, J. (1988). Assessing the occurrence of elaborative inferences: lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399.

Reiss, M.J., Millar, R. & Osborne, J. (1999). Beyond 2000: science/biology education for the future. *Journal of Biological Education, 33*(2), 68-70. doi:10.1080/00219266.1999.9655644

Ritchie, S., Rigano, D., Duane, A. (2008). Writing an ecological mystery in class: merging genres and learning science. *International Journal of Science Education, 30*(2), 143-166. doi: 10.1080/09500690601161783

Rudge, D.W. (2000). Does being wrong make Kettlewell wrong for science teaching? *Journal of Biological Education, 35*(1), 5-11.

Rudge, D.W. (2004) The Mystery Phenomenon: Lesson Plans. In D. Metz (Ed.) *Proceedings of the Seventh International History, Philosophy and Science Teaching Group Meeting* (pp 773-811). IHPST: Winnipeg, Canada.

Rudge, D.W., & Fulford, J.M. (2011). The role of visual imagery in textbook portrayals of industrial melanism. In Seroglou, F., Koulountzos, V. & Siatras, A. (Eds.) *Science & Culture: Promise, Challenge and Demand*, *Book of Proceedings for the Eleventh International History, Philosophy & Science Teaching (IHPST) and Sixth Greek History, Philosophy and Science Teaching Joint Conference, Aristotle University*, *1-5 July 2011* (pp. 630-637). Thessaloniki, Greece: Epikentro Publications.

Rudge, D.W., Cassidy, D.P., Fulford, J.M. & Howe, E.M. (2014). Changes observed in views of nature of science in a historically based unit. *Science & Education*. *23*(9), 1879-1909.

Rudge, D.W., & Howe, E.M. (2009). An explicit and reflective approach to the use of history to promote understanding of the nature of science. *Science & Education, 18,* 561-580.

Saenz, L. M. & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities: expository versus narrative text. *Remedial and Special Education*, *23*(1), 31-41.

Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. Los Angeles, CA. Sage Publishing.

Salmon, M.A. (2000). *The Aurelian Legacy: British Butterflies and Their Collectors.* Berkeley, CA. The University of California Press.

Schick, K.D., & Toth, N. (1993). *Making silent stones speak: Human evolution and the dawn of technology.* New York: Simon and Schuster.

Shrigley R. L., Koballa T.R. (1989). Anecdotes: what research suggests about their use in the science classroom. *School Science Math*, *89*(4), 293–298

Shutt, R. K. (2009). *Investigating the Social World: The Process and Practice of Research. 6th Ed.* Los Angeles, CA: Pine Forge Press. Sage Publications, Inc.

Sinatra, G.M., Brem, S.K., & Evans, E.M. (2008). Evolution Education Outreach, 1, 189-195. doi:10.1007/s12052-008-0037-8

Smith, M.U. (2010). Current status of research in teaching and learning evolution: II. pedagogical issues. *Science & Education*, *19*, 539-571. doi:10.1007/s11191-009-9216-4

Snook, I., O'Neill, J., Clark, J., O'Neill, A.M., & Openshaw, R. (2009). Invisible learnings? a commentary on John Hattie's book: visible learning: a synthesis of over 800 meta-analyses relating to achievement. *New Zealand Journal of Education Studies, 44*(1), 93.

Stinner, A., McMillan, B.A., Metz, D., Jilek, J.M., & Klassen, S. (2003) The renewal of case studies in science education. *Science & Education 24*617-643.

Summers, G., Decker, T., & Barrow, L. (2007). The treatment of geological time and the history of lige on earth in high school biology textbooks. *American Biology Teacher*, *69*(80), 401-405.

Tashakkori, A., & Teddlie, C. (1998, 2003). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks: Sage

Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: calling for an integrative framework. In Bergman (Ed.), Advances in Mixed Methods Research. Thousand Oaks, CA. Sage Publications.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research*. Thousand Oaks, CA. Sage Publications.

Thorndyke, P.W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, 77-110.

Valentine, J. C., & Cooper, H. (2003). Effect size substantive interpretation guidelines: issues in the interpretation of effect sizes. Washington DC: What Works Clearinghouse.

van Dijk, M.E. & Kattmann, U. (2009). Teaching evolution with historical narratives. *Evolution Education Outreach, 2*, 479-489.

Wilson, E. O. (2002) The Power of Story. *American Educator, 26*(1), 10-11.

Wandersee, J.H. (1990a). Can the history of science help science educators anticipate students' misconceptions? *Journal of Research in Science Teaching, 23*(7), 581-597.

Wandersee, J.H. (1990b). On the value and use of history of science in teaching today's science: constructing historical vignette. In D. Herget (ed.), More history and philosophy of science in science teaching. Tallahassee, FL: Florida State University Press.

Wandersee, J.H. (1992). The historicality: implications for science education research. *Journal of Research in Science Teaching, 29*(4), 423-434.

Wandersee, J.H. & Roach, L.M. (1998). Interactive historical vignettes. *Teaching Science for Understanding: A Human Constructivist View*. 281-306. Academic Press.

Wolfe, M.B. & Mienko, J.A. (2007). Learning and memory of factual content from narrative and expository text. *British Journal of Educational Psychology*, 77, 541-564.