8-2017

Effects of Historical Story Telling on Student Understanding of NOS and Mendelian Genetics

Cody Tyler Williams
Western Michigan University, redwings.13@gmail.com

EFFECTS OF HISTORICAL STORY TELLING ON STUDENT UNDERSTANDING OF NOS AND MENDELIAN GENETICS

by

Cody Tyler Williams

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Mallinson Institute for Science Education
Western Michigan University
August 2017

Doctoral Committee:

David W. Rudge, Ph.D., Chair
Heather L. Petcovic, Ph.D.
Patricia L. Reeves, Ed.D.

EFFECTS OF HISTORICAL STORY TELLING ON STUDENT UNDERSTANDING OF NOS AND MENDELIAN GENETICS

Cody Tyler Williams, Ph.D.

Western Michigan University, 2017

Science education researchers have long advocated the central role of the nature of science (NOS) for our understanding of scientific literacy. NOS is often interpreted narrowly to refer to a host of epistemological claims associated with the process of science and the limitations of scientific knowledge. Despite its importance, practitioners and researchers alike acknowledge that students have difficulty learning NOS and that this in part reflects how difficult it is to teach. Many teachers are uncomfortable with taking time away from traditional science content instruction. One promising method for teaching NOS and science content involves an explicit and reflective approach using historical stories. The purpose of this study is to compare a traditionally taught genetics unit in a non-major introductory biology course, to the same genetics unit taught using a historical story based on Gregor Mendel. A mixed method approach was used to determine whether and how the use of historical stories influences undergraduates' understanding of NOS and genetics content. Particular attention was paid to the explanations students used for their understandings. Intervention and control groups completed the SUSSI instrument and a two-tier genetics instrument pre- and post-instruction. A subset of both groups was also interviewed regarding their responses to both instruments and their experiences in the course.

The SUSSI Likert results showed that students in the intervention group made statistically significant gains in their understanding of the role of imagination and creativity in science. These results indicate that the introduction of historical stories helped participants gain a better understanding of this aspect of NOS. The interviews provided additional support in that participants mentioned historical stories in their explanations for why they changed towards more informed views on SUSSI items related to imagination and creativity. Additionally, students recognized that stories were used in the intervention group without prompting and felt they were helpful for learning about science. The genetics two-tier instrument results showed that participants made more statistically significant gains in their genetics content understanding in the intervention group than the control group. The current study adds to a growing body of literature regarding the use of stories in the science classroom. The results provide support for using historical stories to improve student understanding of NOS as well as more traditional science content. This study suggests further research into the role of stories in science instruction.

ACKNOWLEDGEMENTS

I would like to express deep gratitude to my advisor, Dr. David W. Rudge, for his support throughout my graduate career at WMU. He has worked with me since the beginning of my WMU career and has helped me navigate through many highs and lows along the way. Dr. Rudge cares deeply for all of his students and has always strived to help me make the best possible decisions regarding my research and career. I would also like to thank Dr. Heather L. Petcovic and Dr. Patricia L. Reeves for helping me to fine tune my design and taking the time to review my work.

My research group has been one of the most important supports in getting through my Ph.D. program. They have provided me with weekly accountability, emotional support, and timely advice. In particular, I would like to thank Dr. Janice Fulford and Aekam Barot who have been with me since the beginning of my research at WMU. Dr. Andrew Bentley has played a large role in helping to push me through the long days of the end of my program and I thank him for that. A special thank you also to my transcriptionist Tessa Ziebarth.

My family has been the largest source of emotional support throughout my education. I would like to thank my mother and father, Connie and John, for supporting me throughout my life and for their strong belief in a quality education. I would not have been able to reach this point in my career if not for their encouragement, love, and guidance. Last but certainly not least, I would like to thank my partner Katie Melvin for her unwavering support, love, and patience. I would not have been able to complete this project or my Ph.D. without her.

Cody Tyler Williams

TABLE OF CONTENTS

Table of Contents–Continued

CHAPTER

Table of Contents–Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Science educators have long held that an understanding of the nature of science (NOS) is an important and uncommon outcome of science instruction (Central Association of Science and Mathematics Teachers, 1909; Lederman et al., 2002; AAAS, 2009). An adequate understanding of science goes well beyond familiarity with science content knowledge, such as the fact that matter consists of atoms. Students should also know something about how scientific knowledge is generated, and also its limits. A deeper understanding of science also includes considerations from a broader cultural and historical context (Matthews 1994). Although there is wide agreement among science educators about the importance of NOS, there is no single agreed upon definition.

The multitude of definitions that have been suggested for NOS are illustrated by the frameworks described in a review of the NOS literature by Deng, Tsai, and Chai (2011). One of the most widely used frameworks for studying NOS views, particularly in recent literature, is the multidimensional (MD) framework (Deng et al. 2011). This framework conceptualizes NOS as being based on some agreed upon epistemological characteristics of science, such as the creation and development of theories, the tentativeness of scientific knowledge, and the distinction between a scientific theory and a scientific law (Lederman et al. 2002; McComas 2004; Lederman 2007). Importantly, many authors utilizing the MD framework make a distinction between the epistemological features above and the process of science or scientific inquiry which is defined as "activities related to collecting and analyzing data and drawing conclusions" (Lederman et al., 2014, p. 977). Studies under the multidimensional framework often rely on method triangulation between assessment instruments (both open-ended and close-ended) and interviews. Some researchers have argued that a drawback to the MD framework is that the proposed characteristics of science are too generalized and abstract to be practically useful. Similarly, the generalized nature of characteristics does not properly account for the importance of context in NOS views (Elby & Hammer, 2001; Hodson, 2014).

An alternative framework for studying nature of science views is the argumentative resource (AR) framework (Deng et al. 2011). Researchers working under this framework focus on students' ability to construct and evaluate scientific arguments or explanations. A scientific explanation can be defined as a causal account for how or why a phenomenon occurs that accounts for data such as observations or events (Sandoval, 2003). Allchin (2011) states that instead of students learning particular characteristics of scientific knowledge students should learn how to interact with expert scientists, to distinguish between relevant and irrelevant data, to recognize the limits of new scientific claims, and to handle uncertainty in science. These types of studies often employ observation and interview methods using discourse analysis (Deng et al., 2011). The potential disadvantage of the AR framework is that it was most recently developed and has been used relatively less in the NOS literature. As a result, there have been inconsistencies in the implementation of this framework and the criteria for determining whether an individual has sophisticated NOS have not been fully defined (Deng et al., 2011).

Despite inconsistencies in the application of the Argumentative Resources framework, some research has indicated that the quality of students' scientific arguments or explanations can be an indicator of science content knowledge. Research has shown that students' ability to create explanations is related to their NOS and science content understanding. For example, Zeidler et al. (2002) found that high school and undergraduate students' explanations of their views on socioscientific issues (SSI), such as animal testing, were influenced by their NOS views. Similarly, Khishfe (2012) examined student argumentation skills and NOS views in regard to SSIs. High school students were asked to provide arguments, counterarguments, and rebuttals in response to scenarios involving genetically modified foods and adding fluoride to drinking water. The students were also prompted about their NOS views for both scenarios. Quantitative and qualitative data showed a positive relationship between student NOS views and argumentation skills. In terms of science content knowledge, Sadler and Zeidler (2005) investigated the relationship between undergraduate students' genetics content knowledge and informal reasoning skills. Genetics content was measured using a survey and informal reasoning was measured through student arguments regarding six genetics SSIs. The six SSIs concerned gene therapy and cloning. Results showed that students with high genetics content knowledge had significantly fewer reasoning flaws than those with low genetics content knowledge. Despite the positive results in the studies overviewed here, the literature regarding the relationship between content knowledge and scientific explanations or arguments is quite mixed (e.g. Bell & Lederman, 2003; Sadler & Fowler, 2006). The interaction between NOS understanding, science content knowledge, and explanations will be further explored in the current study (see chapter 3).

Emphasis on NOS has also led to several attempts by researchers to improve students' NOS views (Akerson et al. 2000; Kim & Irving 2010; Rudge et al. 2014). Collectively, these and numerous other studies have highlighted how resistant student NOS views are to change (Lederman & Lederman, 2014). One reason NOS views are difficult to change is science teachers' understanding of NOS. Research on teachers has shown that misconceptions related to NOS also extend to science teachers (Lederman, 2007). Another factor is that students often rigorously hold to previously held misconceptions about NOS topics (Clough 2006). Students' difficulty with NOS is likely the result of implicit experiences with what science is both inside and outside of the classroom. Students receive inaccurate implicit messages through cookbook laboratory activities, textbooks, and media reports. These implicit messages lead to students developing deeply held misconceptions regarding NOS, which are difficult to change, even through explicit instruction (Clough 2006).

Multiple researchers have advocated and provided empirical support for using an *explicit and reflective* approach to teaching NOS (Abd-El-Khalick & Lederman 2000; Khishfe & Abd-El-Khalick, 2002; Rudge & Howe, 2009). *Explicit* in this case refers to planned instructional practices that allow for NOS aspects to be openly covered in class. *Reflective* refers to students having the opportunity to come to their own conclusions about NOS aspects and not just repeating what the instructor tells them (Akerson et al. 2000). This approach has its origins in two closely related theoretical frameworks. One theoretical basis for this approach can be found in constructivism, which says that students need to incorporate new understandings into previously learned constructs (Ausubel 1960; Matthews 1997). This means that the instructor must be aware that students have previously held beliefs about NOS (Clough 2006). Constructivism also holds that learning is an active interaction with new material. Students must be given opportunities to incorporate new information and rearrange their existing constructs.

A second theoretical basis for the *explicit and reflective* approach can be found within conceptual change theory (Posner et al. 1982; Duit and Treagust 2003). Appleton (1997) provides a useful empirically based framework for relating conceptual change to NOS instruction (Figure 1). The basis of the framework is that learners seek a "best fit" understanding of a given concept, which minimizes cognitive conflict between new information and previously held understandings. Appleton proposes three pathways that stem from learners' conflict between previous and new knowledge. The first pathway involves learners exiting instruction with an understanding of a given concept that they believe fits with their previous knowledge but does not follow the accepted understanding of the concept. The second pathway says that learners will see an approximate fit of new information with previous understandings. The learners may then see the approximate fit as being close enough and exit instruction or may seek additional information and reexamine their ideas. For the third pathway, learners acknowledge an incomplete fit of new knowledge or cognitive conflict. Students use the new knowledge and new ways of thinking to create a better fit between previous understandings and the new knowledge.

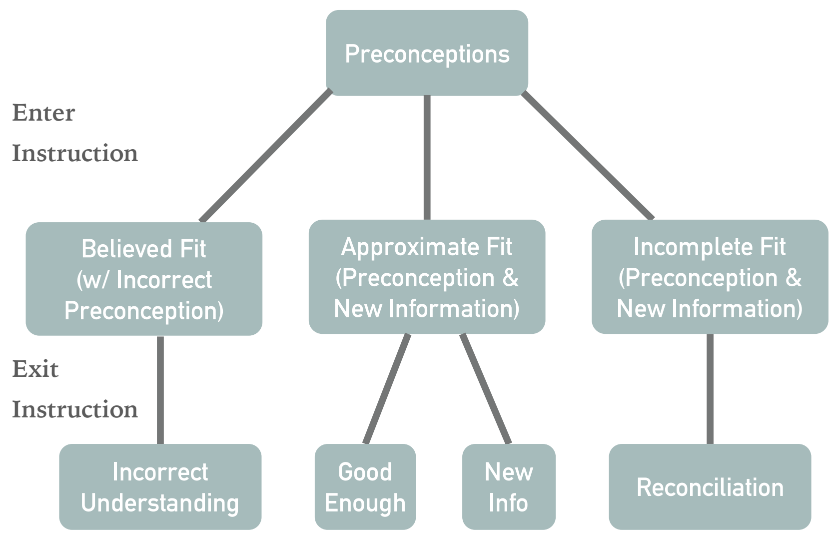


Figure 1. Student learning pathways based on Appleton's (1997) model of conceptual change.

Appleton's framework has implications for NOS instruction in that students' typical encounters with NOS concepts often reinforce previously held misconceptions. Common textbook accounts, lectures, and "cookbook style" activities present students with a portrayal of science that aligns with previous misconceptions (Clough 2006). Even if an instructor designs a lesson to address NOS, students may exit instruction with their misconceptions intact. People naturally interpret new information through the lens of their previous understandings. As a result, they will likely look for aspects of new information that align with previous knowledge. Additionally, they may ignore information that is contrary to their previous conceptions and even modify new information to fit these conceptions (Clough 2006). The above reasons are likely a major contributing factor for why previous research has supported the use of explicit approaches to teaching NOS instead of implicit approaches (Akerson et al. 2000; Kim & Irving 2010; Rudge et al. 2014).

Active engagement with historical understandings of science can help overcome the tendencies of learners to align new learning with previous knowledge. Using the history of science (HOS) for teaching NOS has been advocated by several researchers (Matthews 1994; McComas 2010; Rudge et al. 2014). The HOS is promising because it provides a highly contextualized approach to teaching NOS. Context introduces a human element to NOS instruction. More importantly, it allows for intertwining NOS instruction with science content. Historical accounts of scientists are necessarily tightly bound with science content (Clough 2006). Additionally, instruction informed by the HOS may allow students to compare their understanding with current scientific ideas, in an environment that legitimizes student thinking (Monk and Osborne 1997).

The practical realities of the classroom emphasize the importance of melding NOS instruction with science content. Science instructors often view NOS instruction as wasting instructional time, particularly decontextualized teaching approaches. Decontextualized approaches (e.g. silver box activity, mystery tube activity) can be problematic for instructors and students because they require the instructor to explain the similarities between an abstract activity and science. Instructors can see these decontextualized activities as taking time away from science content. Students can develop multiple views related to NOS, personally held views and those that are appropriate for school settings (Clough 2006). This is not to say there is not value in decontextualized approaches but rather to illustrate the importance of historical or contextualized approaches. Decontextualized approaches provide students with a way of exploring NOS concepts without the added complexity of science content. As a result, they are useful for providing a basic understanding of NOS concepts prior to exposing students to more contextualized examples.

Historical instructional approaches show promise for providing students with contextualized learning experiences for both NOS and science content more broadly. However, there is a lack of a shared operational definition for the use of the HOS in NOS research. The meaning of HOS varies from study to study, which likely has contributed to the mixed results produced by these studies. Some of the studies included in chapter 2 illustrate the issue with

HOS clearly. For example, a study by Abd-El-Khalick and Lederman (2000) involved the evaluation of entire courses based on the HOS. Another study by Kim and Irving (2010) was in direct contrast in that it focused on a single genetics unit. Until there is a clearer understanding of what is meant by researchers when they refer to the HOS, progress will be difficult. Historical stories are a possible solution to the difficulties with operationalizing HOS.

More specific advantages to using the HOS for NOS and science content instruction can be made apparent by examining a specific science discipline. Genetics is one discipline that can particularly benefit from HOS. Gericke and Smith (2014) point out that the field of genetics historically has developed several explanatory models including classical genetics, molecular genetics, and genomics. The difficulties students often have in genetics typically mirror a classical view of genetics. For example, many students have the misconception that genes are the only factor in the development of phenotypic traits. As a result, researchers such as Gericke and Hagberg (2007) have advocated teaching students about historical explanatory models in genetics. It is thought that these models could help students to clarify their own thinking and show them that there is no one "correct" model waiting to be discovered (Gericke & Smith 2014). Despite the advantages, teachers use HOS sparingly. This may reflect many factors including lack of teacher knowledge of the HOS and the practical realities of the classroom such as limited time and the need to teach to formal exams (Monk and Osborne 1997).

Historical stories represent a contextualized approach to teaching NOS that has been advocated for by some science educators (Metz et al., 2007; Klassen, 2009; Hadzigeorgiou et al., 2012). Stories are thought to engage learners and provide a reason for knowing about science content. They can also introduce a human element to science instruction that is often lacking from traditional instructional approaches (Klassen, 2014). Additionally, well designed stories can produce emotional responses in the reader or listener. Neuroscience research supports the importance of emotion for rational thought and for improving focus (Damasio, 1994; Howard, 2000).

The few empirical studies done in science education related to the use of stories have shown that science stories can positively influence NOS views (Fulford, 2016; Hadzigeorgiou et al., 2012; Klassen, 2009). These studies also indicated that the stories should present science concepts together with the human elements of science such as the accomplishments of historical scientists. In order to use science stories for teaching NOS, it is also important that the stories accurately portray NOS concepts. Ensuring accurate portrayals of NOS also emphasizes the importance of using the history of science as it provides rich background and context for creating stories that address NOS (Klassen & Klassen, 2014).

Despite the promise of using stories in science education, there has been little research done into their utility by science educators. There are two main potential reasons for the lack of research into stories. One of the main issues is that until recently there has not been a consistent theoretical framework for creating science stories (Metz et al., 2007, Klassen, 2009; Klassen & Klassen, 2014). Without a theoretical framework, it is difficult for researchers to design studies that are comparable to each other. This has resulted in a limited and fragmented literature where various researchers have introduced different interpretations of stories. Not being able to make comparisons between studies makes it difficult to determine the effectiveness of stories as an instructional approach and to plan future research into stories. The other likely reason for the lack of story research in science education is related to practical concerns. Writing quality science stories that can address both science content and NOS concepts in a way that is engaging for students is a difficult task. It requires technical and creative skills for which many science education researchers do not have formal training. Fortunately, there has recently been significant progress made in creating a theoretical framework for creating science stories. This has resulted in limited empirical research that will be considered in chapter 2.

As outlined above, students and teachers have difficulties with NOS concepts despite extensive attention from researchers. The HOS has long been advocated as an approach to improving NOS views. However, empirical studies related to using HOS to teach NOS have been inconsistent. A primary reason for inconsistencies in HOS empirical studies may be the variable manner that HOS has been implemented in these studies. Recently developed theoretical frameworks for creating historical science stories may provide a path towards a more standardized approach to using HOS in the science classroom. These ideas will be explored further throughout the literature review included in chapter 2.

CHAPTER II

LITERATURE REVIEW

The previous chapter outlines the problem that NOS views have been difficult to improve despite significant attention from science educators. Additionally, the history of science has shown promise for improving NOS views but has been inconsistently used in the literature leading to mixed results. The current chapter delves further into these issues through reviewing the literature and provides support for using historical interventions based around narrative structure to improve NOS views. Klassen & Klassen (2014) note that historical stories are well suited for promoting student interest and engagement with science content including NOS. Yet historical stories remain a largely untapped resource in the science education literature (Klassen & Klassen, 2014). The literature review below covers three areas of research. The first section is an overview of NOS research. The next section establishes that the history of science is a viable method for teaching students about NOS. The third section of the review provides a comprehensive overview of the use of stories in science education.

Teaching and Learning of NOS

The studies reviewed in this section (Table 1) reveal several insights regarding research on the teaching and learning of NOS. Each of the articles is representative of different approaches to NOS research. In addition, instruments for assessing NOS understanding are also considered. For their study, Khishfe and Abd-El-Khalick (2002) focused on the tentative, empirical, and creative and imaginative NOS. The approach taken in this study is consistent with the multidimensional (MD) approach described by Deng et al. (2011), in which NOS views are evaluated in terms of agreement with general statements regarding NOS. In addition to being an example of the MD approach, the study also provides a rigorous comparison of the explicit and reflective and implicit approaches to teaching NOS. The quality design of the study using equivalent experimental and control groups offers compelling evidence for using an explicit approach over an implicit approach. While the Khishfe and Abd-El-Khalick (2002) study focused on 6th grade students, other recent studies have found positive results using an explicit approach (Howe & Rudge, 2005; Hanuscin et al., 2006; Rudge et al., 2014). In addition, recent reviews of NOS research have all concluded that an explicit approach should be used (Lederman, 2007; Deng et al., 2011; Hodson, 2014).

Table 1

Articles Reviewed Related to the Teaching and Learning of NOS

Article Title (Author, Year)
Influence of Explicit and Reflective versus Implicit Inquiry-Oriented Instruction on Sixth Graders' Views of Nature of Science (Khishfe & Abd-El-Khalick, 2002)
Designing and Evaluating Short Teaching Interventions About the Epistemology of Science in High School Classrooms (Leach et al., 2003)
Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories (Tao, 2003)

Leach et al. (2003) and Tao (2003) are examples of studies that have taken a contextualized approach to studying student NOS views. Leach et al. (2003) used a short-term historically based intervention that explicitly addressed NOS concepts. The intervention consisted of a single lesson. Despite validity and reliability issues with the probe used to measure NOS views, the authors found positive changes in student NOS views after the intervention lesson. The study indicates that short-term interventions can be effective in improving student NOS views. This is supported by other studies that have also found positive results from relatively short-term interventions (i.e. Rudge et al., 2014). Tao (2003) used historically based stories to improve students' views of several NOS tenets consistent with the multidimensional framework (MD) view of NOS. The author used an implicit approach to teaching NOS and primarily expected students to improve their understanding by being exposed to the stories. Tao (2003) found modest differences in students' NOS views after the stories, further reinforcing that researchers and instructors should use explicit approaches to NOS. This being said, another finding from this study was that the stories resonated with students. The Tao (2003) study calls for further research into the use of stories for teaching NOS. Combining the use of stories with an explicit approach to teaching NOS may have resulted in greater improvement in students' NOS views. Stories could be an important tool in improving NOS views as they interesting to students, as evidenced by Tao (2003), and they can contextualize NOS concepts. These ideas are further explored in the third section of the literature review below.

An important consideration when conducting NOS research is available instruments for assessing NOS views. Many of the instruments from the 1950s, 1960s, and 1970s are no longer viable for current NOS research (Lederman and Lederman, 2014). The majority of these instruments have issues with validity and reliability and or outdated NOS content. There are two primary instruments that are currently commonly used in NOS research. The instruments are based on the multidimensional framework for NOS. There are currently no valid and reliable instruments available that are based on the Argumentative Resource framework (Deng, 2011). Both of the commonly used instruments have been extensively tested for validity and reliability. These instruments are the VNOS (Lederman et al., 2002) and the SUSSI (Liang et al., 2008). Deciding which instrument should be used depends primarily on the research questions being answered. The primarily qualitative nature of the data collection and analysis involved with administering the VNOS makes it more suitable to smaller scale studies. Studies that are more exploratory and require rich descriptions of individual NOS views are likely better oriented toward using the VNOS instruments. Larger scale studies where researchers are interested in making generalizable claims using inferential statistics are more conducive toward the SUSSI. The SUSSI has some additional flexibility in that it also includes qualitative data collection in the form of open response items. SUSSI open response questions provide a built-in check on the quantitative data and can provide additional nuance regarding participant NOS views.

The foregoing review establishes that the explicit and reflective approach to teaching NOS is preferred by most researchers instead of an implicit approach (Khisfe & Abd-El-Khalick, 2002; Lederman, 2007; Deng et al., 2011). Despite agreement on the general approach to teaching NOS, there is disagreement about whether NOS instruction should be contextualized. Clough (2006) argues that there is a role for both contextualized and decontextualized examples in NOS instruction. In general, decontextualized examples can be useful for introducing individuals to NOS concepts of which they are unfamiliar. Contextualized examples on the other hand can help to provide deeper understandings once some familiarity with an NOS concept has been achieved. Context introduces a human element to NOS instruction. More importantly, it allows for intertwining NOS instruction with science content (Clough, 2006). The review also showed that there are two main instruments that are currently used often in NOS research, the VNOS and the SUSSI.

Implementing the HOS to Affect NOS Views

One method that has been advocated for teaching NOS in a contextualized manner is through the use of the history of science (HOS) (Matthews, 1994; McComas, 2010; Rudge et al., 2014). Several potential benefits for using historical approaches were put forward by Matthews (1994) including the addition of a human element to science instruction that is often missing, providing students with opportunities to practice their critical thinking skills, and providing teachers openings to address student misconceptions since they are often similar to those of historical scientists. Despite the advantages, teachers use HOS sparingly. This may reflect many factors including lack of teacher knowledge of the HOS and or NOS and the practical realities of the classroom such as limited time and the need to teach to formal exams. Another likely contributing factor is a lack of quality historical materials available for teachers (Monk and Osborne 1997).

A wide variety of approaches have been taken to study the use of the history of science for improving NOS views as demonstrated by the papers reviewed in this section (Table 2).

Table 2

Articles Reviewed Related to Implementing the HOS to Affect NOS Views

Article Title (Author, Year)
Placing the History and Philosophy of Science on the Curriculum: A Model for the Development of Pedagogy (Monk & Osborne, 1997)
The Influence of History of Science Courses on Students' Views of Nature of Science (Abd-El-Khalick & Lederman, 2000)
Promoting Preservice Chemistry Teachers' Understanding about the Nature of Science through History (Lin & Chen, 2002)
Recapitulating the History of Sickle-Cell Anemia Research (Howe & Rudge, 2005)
History of Science as an Instructional Context: Student Learning in Genetics and Nature of Science (Kim & Irving, 2010)

Several authors have made theoretical arguments for including the history of science in science instruction (e.g. Matthews, 1994; Rudge & Howe, 2009). The paper by Monk and Osborne (1997) serves as a highly cited and influential example of the theoretical arguments for the history of science. In addition, there have been many different kinds of interventions created using the history of science for the purposes of improving NOS views. The pieces by Abd-El-Khalick and Lederman (2000) and Lin and Chen (2002) are examples of course length interventions. While Howe and Rudge (2005) and Kim and Irving (2010) examine interventions that span individual units within a course. Reviewing the set of papers included in this section has led to several insights regarding the use of the history of science for improving NOS views and the needed future steps in this area of research.

Monk and Osborne (1997) argued that the history of science can be helpful for improving NOS views because it can provide accurate depictions of the scientific process. Additionally, student explanations often align with the explanations of historical scientists which can allow for legitimizing student ideas. To capitalize on the potential advantages of using the history of science, Monk and Osborne propose a model for incorporating the history of science. Importantly, the model shows one way in which constructivist principles, in the form of conceptual change, can be combined with the history of science in the science classroom. The model aligns with conceptual change by calling for eliciting students' prior beliefs, presenting a historical episode, and tasking students with comparing their views to historical views. However, there are also some problematic aspects of the Monk and Osborne model that need to be considered. Rudge and Howe (2009) criticize the Monk and Osborne model for relying upon students implicitly learning about NOS through exposure to historical materials. They point out that having students listen to a lecture or read materials about historical perspectives without students actively considering their currently held beliefs is likely to be a passive learning experience. As a result, students would be more likely to view historically based instruction as an add-on and a waste of time. Rudge and Howe's view of the model is supported by numerous studies showing that an implicit approach is not as effective for teaching NOS when compared to explicit approaches (Abd-El-Khalick and Lederman, 2000, Tao, 2003; Deng et al., 2011). A more appropriate approach to incorporating the history of science, according to Rudge and Howe

(2009), is to explicitly elicit students' NOS views by asking them to consider the NOS concepts presented in a historical episode.

Abd-El-Khalick and Lederman (2000) examined the effects of history of science courses on student NOS views. They also compared a course that used an explicit and reflective approach to one that used an implicit approach. While the authors found support for the explicit and reflective approach, they did not find support for using the history of science. The lack of support for the history of science may have been the result of evaluating the effects of full courses. Choosing an entire course as your unit of analysis potentially adds considerable noise to the interpretation of the results. In their study, Abd-El-Khalick and Lederman (2000) looked at courses that each had distinct learning objectives, instructors, and instructional formats. Some of the courses taught NOS principles explicitly while one of the courses did so implicitly. Differences between the courses make comparisons difficult and limit the conclusions that can be drawn. As the authors note, there is more research needed to determine the effectiveness of HOS for improving NOS views. Interestingly, Abd-El-Khalick and Lederman (2000) also concluded that conceptual change should be used with the history of science in the science classroom in agreement with research discussed previously (Monk & Osborne, 1997; Rudge & Howe, 2009).

Further support for conducting additional research on using the history of science for improving NOS views comes from the study by Lin and Chen (2002). Despite concerns regarding the literature review, instrument validity, and the instructional approach used in the intervention the authors found some positive results using the history of science. Lin and Chen (2002) found that after a historically based intervention some students were better able to articulate their NOS views using examples from the history of science. The promise of these results is enhanced by the quality of the overall research design as the authors used a true comparison group and a mixed methods design.

Studies by Howe and Rudge (2005) and Kim and Irving (2010) give some ideas regarding where research related to the history of science and NOS should go in the future. Howe and Rudge's (2005) paper is an example of an interrupted narrative approach to incorporating the history of science. One issue with this study is that it only includes limited empirical evidence. However, Rudge et al. (2014) used a similar instructional approach in a rigorous empirical study. The Rudge et al. (2014) study used the VNOS instrument and follow up interviews to study an intervention related to Industrial Melanism and the work of H.B. Kettlewell. The results of this study showed that student's NOS views improved after the intervention. While both Howe and Rudge (2005) and Rudge et al. (2014) show the promise of the interrupted narrative approach, they do not provide a standardized method for incorporating the history of science. Future research should address standardizing how history is implemented in order to improve the comparability of studies in this research area. Additional discussion on this point is included in section 3 of the review.

Kim and Irving's (2010) study presents some evidence that history of science can be used in the context of genetics to improve NOS views without sacrificing student learning of science content. This being said, the instrument used in this study was not validated and there was not a sufficient argument given for its use. As a result, there is some question regarding the validity of the results. The study does still further reinforce that the history of science shows promise for teaching the history of science. It also serves to begin to address one of the major concerns of science teachers regarding instruction involving the history of science, namely that history takes away from time spent on traditional science content. Kim and Irving's (2010) study is also evidence that genetics should be considered as a context for future research regarding the use of the history of science and NOS.

Stories in Science Education

Jerome Bruner (1986) stated that there is little known about creating quality science stories for the classroom and recommended that researchers should attempt to address this gap. Despite this call to action, there has been limited research published regarding stories in the science education literature (Klassen & Klassen, 2014). Only 18 articles have been published related to science stories since 1996. The majority of these are theoretical in nature with only two being empirical. The theoretical studies have primarily focused on discussing features of stories that should make them well suited for teaching science content or NOS. Some have also provided exemplars of stories actually being used in the classroom but lacked rigorous empirical data supporting their utility. The papers included in this section of the review were taken from a comprehensive review of articles related to the use of science stories done by Klassen and Klassen (2014) (Table 3). Klassen and Klassen's (2014) review included all peer reviewed published research related specifically to science stories. The articles chosen for the current review were deemed to be the most important because they provided important definitions, a theoretical framework, and empirical evidence supporting science stories. The first two articles reviewed in this section are recent examples of theoretical articles that have been published related to science stories. Hadzigeorgiou et al. (2012) is the most recent and complete empirical study available in the science education literature. Klassen (2009) is a primarily theoretical paper that includes a small amount of empirical data. It was chosen for the review because it provides a framework for designing science stories for future research and instruction.

Table 3

Articles Reviewed Related to Stories in Science Education

Article Title (Author, Year)
Encouraging a ‘‘Romantic Understanding’’ of Science: The Effect of the Nikola Tesla Story (Hadzigeorgiou et al., 2012)
The Construction and Analysis of a Science Story: A Proposed Methodology (Klassen, 2009)
Building a Foundation for the Use of Historical Narratives (Metz et al., 2007)
The Relation of Story Structure to a Model of Conceptual Change in Science Learning (Klassen, 2010)

The lack of research into science stories by science educators is also reflected by Metz et al. (2007) finding that there was a lack of agreement regarding definitions of basic terms such as “story”, “storyline”, and “narrative” in science education. In providing these definitions, Metz et al. (2007) took an important first step toward moving story research forward in science education. A second major step to set the stage for future research was taken by Klassen (2010). The main contribution of this study is that it provides a theoretical basis for stories that should be highly relevant to many science education researchers. Klassen (2010) shows a clear connection between a minimal narrative structure and an interpretation of conceptual change theory. This being said, the study is problematic in that it does not fully explain how stories are able to motivate individuals to learn. Lacking this explanation is particularly problematic since motivation to learn is a key component of conceptual change. As a result, there is a need for further research into establishing conceptual change theory as a fit for science stories and stories more generally.

Hadzigeorgiou et al. (2012) represents the only primarily empirical study published recently in a science education journal. The theoretical studies discussed above seem to have begun to move some researchers toward investigating the utility of science stories. The authors found positive results in terms of science content understanding and emotional involvement with science after exposing students to a unit based around a historical story. However, there were also concerns related to the nature of the historical story used for the study in that it overemphasized the romantic aspects of the story. The use of romanticized history along with problems in the data collection related to reliability and validity call for more empirical to be done related to science stories.

Guidance for conducting future work related to science stories can be found in Klassen (2009) which provides a framework for creating and evaluating the quality of science stories. The framework consists of ten elements that should be included in narratives (Table 4). Each element is based in narrative theory. In creating the framework, the teaching and learning of science was also considered. The Klassen (2009) framework also represents a notable contrast with the approach taken by Hadzigeorgiou (2012) in that they specifically recommend against romanticizing historical science stories. Klassen (2009) points out that romanticizing historical science stories can give a distorted view of NOS. This point is consistent with the views of other science educators such as Allchin (2003, 2011). Despite the importance of the Klassen framework, there were also weaknesses associated with the study. Namely there is only relatively weak empirical data presented in support of the framework. The authors provide some data on questions that were asked by students in response to a story created using the framework. Data was also included on what the students believed was the primary purpose of the story. These data may say something about student's engagement and understanding of the story. However, they do not provide a rigorous measure of student science content or NOS understanding. Obtaining data on both the science content and NOS understanding is critical to understanding the importance of the framework.

Fulford (2016) represents the best empirical data collected on the Klassen (2009) framework to date. For the Fulford (2016) study, data was collected on undergraduate students' understanding of evolution content before and after a historical story regarding the work of

H.B.D. Kettlewell. An experimental and control treatment were used. Students in the experimental treatment were taught using a historical story using all ten of Klassen's suggested narrative elements (Table 4). The control treatment used a similar story regarding Kettlewell that did not use all of Klassen's elements. Results showed that students who received the version of the story that included Klassen's elements had significantly fewer evolution misconceptions than those who received the version of the story without the elements. Despite these promising results, Fulford (2016) calls for future research on the Klassen framework that involves comparisons to instruction with minimal story elements. The current study will build off the encouraging findings and recommendations of the Fulford (2016) study. A primary goal of the current study will be to further establish the utility of historical stories for the science classroom. As discussed previously, an important factor in determining the importance of history for science teaching is to provide more standardized approaches for using history in research studies. Therefore, the current study will seek to provide further evidence whether the Klassen (2009) framework is suitable as a standard approach to employing to historical stories.

Purpose of the Study

The purpose of the current study is to build off of the recent science education literature related to historical stories and empirically test Klassen's (2009) stories framework. Klassen notes ten narrative elements that should be included in a historical story intended for the classroom (Table 4). The study compares a traditionally taught genetics unit in a non-major introductory biology course to the same genetics unit taught using a historical story adapted from B. Williams et al. (2010). The historical story used in this study is based on the work and life of Gregor Mendel. The story is used to introduce students to Mendelian genetics and basic inheritance patterns. A mixed method approach will be used to determine whether the use of a historical story developed using Klassen's (2009) narrative elements has an effect on student understanding of NOS and genetics content. The quantitative portion of the study will involve pre- and post-assessments for both NOS and genetics content and the qualitative portion will include semi-structured interviews. The interviews will focus on the explanations that students give for their responses on NOS and genetics content assessments. The research questions guiding this study are:

Q1. What differences in NOS understandings are revealed from pre- to post-instruction based on participants' SUSSI Likert and open response scores in both the traditional and historical story groups?

Q2. What differences in Mendelian genetics understandings are revealed from pre- to post-instruction based on participants' two-tier survey scores in both the traditional and historical story groups?

Q3. What types of explanations do participants use for changes in their SUSSI Likert responses from pre- to post-instruction, as revealed by the interviews, in both the traditional and historical story groups?

Q3a. What connections exist between the participants' quantitative SUSSI results and their explanations during the interviews?

Q4. What awareness do the participants have regarding the use of story and its associated narrative elements?

Table 4

Narrative Elements from Klassen (2009)

Narrative Element
Event-tokens
• Narratives are composed of a series chronological related events. • The events take place in particular settings and result in changes in state for the involved characters.
Narrator
• A participant or observer relates the story to the reader. • Provides the purpose of the story. • Determines what events are shared with the reader and in what order.
Narrative appetite
• Story told in a manner that increases the readers need to know what will happen next. • Examples of how to increase reader interest include introducing foreshadowing and or suspense.
Past time
• Narrative describes events that occurred in the past. • Does not require that events are presented in chronological order
Structure
• Overall narrative consists of some opening situation, complications, and a resolution. • Similarly, the overall structure of the story contains some number of “minimal stories” represented by initial state -> event -> final state.
Agency
• Narratives have characters that are moral agents (usually human beings). • Decisions made by characters have consequences for those characters.
Purpose
• Provide the reader with an improved understanding of their world. • Many narratives also seek to make the reader feel empathy for the characters.
Role of the Reader
• Narratives are developed with certain expectations in mind for the reader. • Example expectations include the reader recognizing and interpreting the genre of the story, wanting to know what will happen next, developing empathy, and generating questions.
Effect of the Untold
• Some details are intentionally left out of the narrative. • Encourages reader to attempt to fill missing details themselves or generate questions. • Narrative theorists have suggested that this process of considering missing details improves reader engagement.
Irony
• Many times narratives have endings that are unexpected for the reader. • Quality narratives do not require ironic endings. • Irony element is less essential element relative to others.

CHAPTER III

METHODS

Overview

The current chapter describes the methodological approach used to answer the above stated research questions. It begins with a summary of the philosophical stance underpinning the study and two theoretical frameworks that the study draws upon. The chapter continues with a discussion of the mixed method approach used including data collection, analysis, and integration procedures. Finally, issues related to mixed method validity and reliability are considered as well as potential limitations to the study.

Philosophical underpinning. There is a long running debate between purists of both quantitative and qualitative research traditions regarding which tradition is best suited towards answering research questions. Those that advocate for quantitative research typically subscribe to some form of positivism. A quantitative purist following positivism would hold that science including the social sciences follows a process of confirmation and falsification using objective methods. In the view of these researchers, it is possible and preferable to make generalizations free of context and time (Johnson & Onwuegbuzie, 2004). In contrast, qualitative purists often argue for research informed by some combination of constructivism, relativism, idealism, humanism, and hermeneutics. These researchers argue for the existence of multiple realities. As a result, generalizations must be restricted by time and context. Additionally, fully separating causes and effects is not possible and the only source for information regarding a particular reality is the individual experiencing that reality. Qualitative purists prefer a writing style that emphasizes rich descriptions that distill participants' experiences as directly as possible, often using more informal language (Johnson & Onwuegbuzie, 2004).

The debate between these two research camps seems to force researchers to choose one or the other approach when designing a study and not both. This can be problematic depending on the particular research questions since there are strengths and weaknesses to both approaches. For example, a strength of quantitative research is that it is capable of generalizing research results based on replication. A weakness of quantitative research is that it can miss important phenomena operating in a given context because of its focus on theory testing as opposed to theory generation. On the qualitative side, a strength is that it can allow the researcher to describe individuals' experiences which allows for documenting the effects of local contexts. However, qualitative research is often more time-demanding and usually does not allow for generalizations (Johnson & Onwuegbuzie, 2004). Restricting one's self to one or the other research tradition means accepting weaknesses or limitations that may be avoided with more methodological flexibility.

Taking into account the above discussion regarding quantitative and qualitative research traditions, the current study will attempt to maximize the strengths of both traditions while minimizing the weaknesses through using a mixed method approach. As a result, the methodological approaches used in this study will be informed by pragmatism. The philosophical tradition of pragmatism is an attractive fit for mixed method research because it attempts to dispatch of making methodological decisions based on strict adherence to a particular philosophical position. One of the primary tenets of pragmatism that can be seen across the works of classical pragmatists including William James, Charles Peirce, and John Dewey is that the value of an idea or theory should be determined through evaluating the practical consequences of adopting that idea or theory. If the consequences of adoption for a given idea or theory are more positive or desirable for society than the alternative, then it should be adopted. The implication for educational research is that methodological approaches should be chosen that have the best opportunity for answering the research questions posed (Johnson & Onwuegbuzie, 2004; Rorty, 2000; Garrison, 1994).

Theoretical frameworks. There are two related theoretical frameworks that inform the current study. These frameworks are constructivism and conceptual changes. As stated in chapter 1, constructivism is the idea that learners incorporate new information into previously learned constructs (Ausubel 1960; Matthews 1997). Some of the implications of constructivism are that instructors need to be aware that students have previously held beliefs regarding any given topic of instruction. Additionally, students need to be given opportunities to reflect upon new learning in order to rearrange their existing constructs (Clough, 2006). For conceptual change, Appleton's (1997) conceptualization is being used which says that learners will seek the "best fit" understanding of a concept that minimizes dissonance between new learning and previously held understandings. The implication of conceptual change is that instructors must strive to actively engage students in their learning and give them the opportunity to actively reflect on newly learned concepts. These theoretical frameworks were chosen because they are consistent with the instructional approach used in the intervention utilized in the current study. They are also aligned with the beliefs and assumptions regarding teaching and learning held by the primary researcher.

NOS framework. The literature reviewed in chapters 1 and 2 showed why historically NOS, as an overall construct, is considered important and how researchers have approached studying the teaching and learning of NOS. Two primary frameworks were identified, the multidimensional (MD) framework and the Argumentative Resource (AR) framework. While the AR framework shows promise, the MD framework has been chosen for this study. The MD framework was selected because the currently available, valid, and reliable instruments for collection of data on NOS views have all been developed from the perspective of this framework. Despite the previously discussed criticisms of the MD framework regarding its generalized and abstract nature, the presence of valid and reliable instruments represents an important advantage. Insights can be gained regarding NOS views without the need for developing new instruments. Additional details regarding the selection of a data collection instrument for NOS views are included below in the Quantitative Data section (pg. 34).

Research design. This study was conducted using a quasi-experimental design with a non-equivalent control group. The study participants consisted of two classes of undergraduate students representing control and intervention groups. The unit of analysis for the study was each individual class of students. The research questions for this study required the use of both quantitative and qualitative data. Several types of data were collected for each class including survey data, open response data, and interviews.

The current study utilized a mixed method approach. Mixed methods can be defined as "the collection or analysis of both quantitative and/or qualitative data in a single study in which the data are collected concurrently or sequentially, are given a priority, and involve the integration of the data at one or more stages in the process of research" (Cresswell et al., 2003, p. 165). This being said there are a multitude of different approaches that can be used under the mixed methods umbrella. To make sense of the overwhelming number of approaches, three mixed methods typologies were reviewed in choosing the design for this study (Leech and Onwuegbuzie, 2009; Onwuegbuzie and Teddlie, 2003; Cresswell et al. 2003).

Onwuegbuzie and Teddlie (2003) provide a relatively simple data analysis typology that offers a choice between Concurrent Mixed Analysis, Parallel Mixed Analysis, and Sequential Mixed Analysis. The main distinction between Concurrent Mixed Analysis and Sequential Mixed Analysis is the timing of the mixing. In Concurrent Mixed Analysis, quantitative and qualitative data are analyzed at the same time. In contrast, either qualitative or quantitative data is analyzed prior to the other data type in Sequential Mixed Analysis. Parallel Mixed Analysis involves considering both data types at the same time but not allowing for any mixing until after the completion of data analysis. Onwuegbuzie and Teddlie's (2003) typology was not viewed as a fit for the current study because it offers little guidance to the researcher regarding how and when to mix quantitative and qualitative data. Additionally, the distinction between Parallel Mixed Analysis and the Concurrent Mixed Analysis is arbitrary and impractical. It seems extremely difficult to conduct data analysis of both quantitative and qualitative data at the same time while not mixing until after the analysis is completed.

Next the Cresswell et al. (2003) typology was considered. Similar to the Onwuegbuzie and Teddlie (2003) typology, little guidance is offered regarding actually conducting a mixed methods study. Out of several different designs described by the authors, the design that seems to best fit the current study is the concurrent triangulation design. A study following the concurrent triangulation design utilizes quantitative and qualitative data that is collected separately. The two data types are then combined during data interpretation. Both data types build off one another to better answer research questions through triangulation. However, the authors do not provide much additional guidance beyond these generalities making the process of applying the design difficult.

The typology selected for the current study comes from Leech and Onwuegbuzie (2009). This typology was selected because the authors provide a more practically useful set of study designs. The typology consists of 8 different designs ranging from partially mixed to fully mixed. Including several more designs than the previous typologies allows the authors to go into more depth and provide more guidance to researchers. For example, Leech and Onwuegbuzie offer recommendations regarding when to mix data across all phases of the study. Additionally, the authors include sample studies for each of the design types, making the process of selecting a design much more streamlined for the reader. The design from the Leech and Onwuegbuzie (2009) typology that best fits the current study is the Fully Mixed Concurrent Dominant Status design. This design calls for mixing data across multiple phases of the study, which may include the research questions, data collection, analysis, and interpretation. Either the quantitative or qualitative data sources are given more weight in the results of the study. This design was deemed the best fit because of the nature of the survey instrument being used for the current study. The survey instrument involves the collection of quantitative and qualitative data simultaneously. The current study gives more weight to the quantitative results. This is not to diminish the importance of the qualitative results. However, a significant portion of the qualitative data was used to complement the quantitative data which necessarily leads to more weight being assigned to the quantitative portion of the study.

Procedure

This study was conducted with the approval of the Human Subjects Institutional Review Board (HSIRB) at Western Michigan University (WMU) (Appendix A). The study was conducted in the context of BIOS 1120 at WMU. Data was collected from one section of BIOS 1120 during both the Fall 2014 and Spring 2016 semesters. A total of ninety-one students participated in the Fall 2014 semester and ninety-four participated in the Spring 2016. Both semesters consisted of 70% female students. The majority of students in both semesters were white. Participants in the spring 2016 semester were significantly older($\bar{x}$=19.3) than Fall 2014($\bar{x}$=18.6) ($p<.05$). Spring 2016 semester participants had significantly fewer high school science courses ($\bar{x}$=3.05) than Fall 2014 ($\bar{x}$=3.48) ($p<.05$). There were no significant differences in the number of college science course, philosophy courses, and final overall grades between the two semesters of participants.

Students recruited in the Fall 2014 semester became the control group and received traditional Mendelian genetics instruction. Instruction in the control group consisted of didactic lectures, homework problems, and readings from the textbook. Students recruited in the Spring 2016 semester became the intervention group. Instruction in the intervention group was similar to the control group with the addition of a historical story. The story was used to introduce the Mendelian genetics unit during the intervention group semester. Students were not made aware of whether they were assigned to the control or intervention treatments.

Participant recruitment and informed consent process. Each BIOS 1120 section had a maximum enrollment of 150 students. The course is an introductory biology course for non-major students that satisfies general education requirements. Two inclusionary criteria were applied to all participants in the study. First, they had to be enrolled in either the Fall 2014 or Spring 2016 section of BIOS 1120. Second, all participants had to complete the pre- and post-assessments for NOS and genetics content. A subset of participants were also recruited for interviews related to their responses on the pre- and post-assessments. Recruitment was completed on the first day of class with an invitation included on the consent form for the study. Those participants who indicated interest in participating in the interviews were contacted through email after the completion of the post-assessments. The recruitment scripts are included in Appendix B. Fifteen students were interviewed in Fall 2014 and twelve in Spring 2016. No exclusionary criteria were used in this study.

Access to students was obtained through the course instructor. The instructor of the course is separate from the primary researcher and had no role in the data collection process. Consent forms were provided to students on the first day of class each term. Students were informed about the study and given important information regarding the study. Information included in the consent form included contact information for the primary researcher, a general purpose for the study, expectations for participation, and potential benefits and risks associated with participating in the study. All of these topics were discussed with the students during class time by the primary researcher. Importantly, students were explicitly reminded that participation in the study would have no effect on their grade outside of potential extra credit associated with participation in the interviews. After the study was explained and students had an opportunity to ask questions, students interested in participating were asked to sign the consent form. The primary researcher collected all signed consent forms and stored them securely in a locked office. Alternatively, students had the opportunity to sign the recruitment form outside of class and return to the primary researcher at a later date. The course instructor had no involvement in recruitment.

Control group instruction. Participants in the control group received traditional genetics instruction with minimal historical content and no stories using the story elements from Klassen (2009). BIOS 1120 consists of three 50-minute class meetings a week throughout the 16-week semester. The course was taught by means of a flipped classroom. Students were expected to view lectures, complete quizzes, and read the textbook outside of class time. During class, students reviewed material or were introduced to new material through questions given using a classroom response system, short writing activities, class discussions, and short lectures. Students were also given the opportunity to ask their professor any questions they may have had regarding the material they were working on outside of class. The genetics unit itself took place over six class periods. Topics covered during the genetics unit included DNA structure, mitosis and meiosis, basic inheritance patterns, and the cell cycle. DNA replication, transcription, and translation were also discussed. Additionally, students completed practice problems related to inheritance patterns. It is important to note that the assigned textbook readings for the genetics unit did mention Gregor Mendel. However, only minimal historical information was included. The textbook authors noted that Mendel conducted his experiments with pea plants in the 1800s, that his experiments were elegant, and that he applied statistics to inheritance research. The authors then quickly moved on to a basic model for how inheritance functions and describing Mendel’s laws without further mention of Mendel’s work or life.

Intervention group instruction. The intervention group received the same genetics instruction as described above for the control group, with the exception of the addition of two historical stories. The first story was about the life and work of Gregor Mendel. The story was presented over two class periods and was adapted for use in class from a story developed by B. Williams et al. (2010). Class time for inheritance practice problems and review were replaced with the Mendel story. As a result, there was no difference in the amount of time spent on the genetics unit between the control and intervention groups. The story was developed with the previously discussed story elements in mind (Clough, 2011; Metz et al., 2007). Examples of how the story elements were incorporated into the story are included below (Table 6). The Mendel story was intended to introduce students to Mendelian genetics and multiple NOS ideas. An *explicit* and *reflective* instructional approach was supported by the story. Several discussion questions were presented throughout the story that encouraged the reader to consider and reflect upon their understanding of NOS concepts. Four of the questions used during the intervention were developed by the authors of the story (Williams et al., 2010). An additional question, added by the current researchers, was also used (Table 5). The story was presented to students by the instructor over the course of two class periods. The instructor read the story to students in segments. The segments were broken up by the discussion questions which students considered in small groups. The instructor also led class discussions regarding the questions once students had a chance to discuss the questions in small groups. Following the Mendel story, students moved into a unit covering Mendelian genetics. Students learned concepts related to Mendelian genetics including monohybrid crosses, Punnett squares, and family pedigrees.

The second story provided an account of the scientists involved in studying Industrial Melanism in England. Throughout this manuscript the Industrial Melanism story will be referred to as the moth story. The moth story is intended to teach students about natural selection and invites students to try to come up with their own explanations for why dark form moths were becoming more prevalent in England after the industrial revolution. This story was presented in the same fashion as the Mendel story and included Klassen's (2009) story elements. It was presented over three class periods in segments read by the instructor with discussion questions interspersed between the segments. During day two of the story a series of questions was presented which had students think about scientific theories, where theories come from, and how they can be compared. The final day of the story described three explanations for the prevalence of dark form moths that were proposed at the time by scientists. Students were asked discussion questions about how scientists decide among competing theories, whether experiments are always necessary to develop scientific knowledge, and what constitutes scientific evidence. These questions address the NOS concepts of science as a social endeavor and scientific methodology respectively. Following the moth story, students moved on to the evolution unit of the course.

Table 5

Questions Included with Mendel Story

Question	NOS concept
1. Explain how Mendel's thinking shows both a gradual progression from prior ideas regarding heredity and also a break from those prior ideas.	Science is tentative but durable.
2. How does Mendel's work illustrate that observation and data analysis is not objective (i.e. scientists ''see'' through the lens of their theoretical commitments)?	Scientists are influenced by their backgrounds (culture, training, etc.).
3. Do you think that Mendel's convictions about how heredity works influenced his observations? Please provide an explanation for your response.	Scientists are influenced by their backgrounds (culture, training, etc.).
4. Many students today choose not to pursue science careers, thinking that science does not require creativity. How does Mendel's original idea, approach to testing that idea, and his analysis of data illustrate that science is a creative endeavor?	Creativity is used throughout the scientific process.
5. Consider that Mendel's ideas involved "factors" for particular traits, and the application of mathematics and probability to biological systems. Why might scientists in Mendel's time have found these ideas difficult to accept?	Science is a social endeavor requiring communication and debate among a community of researchers.

Note. Four of the questions (Items 1, 2, 4, 5) included with the story were created by Williams et al. (2010).

Table 6

Story Elements Included in the Mendel Story

Narrative Characteristic	Example from Mendel Story
Event-tokens - Narratives are composed of a series chronological related events.	Mendel being recognized as gifted student→ being sent to boarding school→earning top grades and gaining self-discipline/broken nerves
The Narrator - The narrator, either a participant in the story or an observer, determines the point and purpose of the story and selects the events and their sequence.	The narrator of the story is an observer. This is evident in the story beginning with the event of the visit by the seed salesman. The story then flashes back to the beginning of Mendel's life and tells the story in relative chronological order.
Narrative Appetite- Story told in a manner that increases the readers need to know what will happen next.	The author foreshadows at the beginning of the story when they state "but such pressure burdened him with broken nerves that would haunt him for the rest of his life."
Past Time- Narrative describes events that occurred in the past.	The story all takes place primarily during the 19^{th} century with brief flashes back to the 16^{th} and 17^{th} century.
The Structure- The overall structure of the story contains some number of "minimal stories" represented by initial state -> event -> final state.	Several scientists interested in heredity and hybrids are described →these scientists in part inspire Mendel to conduct his experiments → the result is Mendel's insights about heredity.
Agency- Narratives have characters that are moral agents (usually human beings).	Mendel needs to decide whether to attend university. The family farm is on the verge of collapse after an injury to his father but Mendel decides to go to university. The result is that Mendel's sister and her husband take over the farm. The consequences are not entirely clear other than Mendel received substantial money.
The Purpose - Provide the reader with an improved understanding of their world. Many narratives also seek to make the reader feel empathy for the characters.	The purpose of the story is to give the reader a better understanding of how science works through the example of Mendel. An example of raising empathy is the description of Mendel's inability to pass the teaching certification exam.
The Role of the Reader or Listener- Narratives are developed with certain expectations in mind for the reader.	The story is developed with the expectation that the reader will understand that it is a historical story designed to give the reader a better understanding of how science works through examples from Mendel's life and work.
The Effect of the Untold - Some details are intentionally left out of the narrative to engage reader.	A short story like the Mendel story necessarily leaves out many details of Mendel's life.
Irony- Many times narratives have endings that are unexpected for the reader.	Mendel did this work that is acknowledged as being important throughout the story by the author/narrator. However, the work was not recognized by the scientific community for over 30 years, long after the death of Mendel.

Intervention rationale. The B. Williams et al. (2010) Mendel story was selected for use in this study for multiple reasons. One of the research goals of this study was to determine whether historical stories are effective in teaching students introductory genetics content. The Mendel story is consistent with this goal in that it describes a historical episode relevant to genetics. Another important consideration is that the story includes ten story elements suggested as being important by science education researchers (Klassen et al., 2009; Metz et al., 2007) (Table 5). Finally, consistent with the science education literature on stories more generally, there is little empirical evidence regarding the effectiveness of the Mendel story for improving student NOS views or genetics content knowledge. All of these reasons together make the B. Williams et al. (2010) story ideal for study.

Data Collection

Both quantitative and qualitative data were collected for this study. Quantitative and qualitative survey data were collected at the beginning and end of both the Fall 2014 and Spring 2016 semesters. In addition, qualitative interview data was collected at the end of both semesters. As a result, quantitative data was collected at the same time as qualitative data.

Quantitative data. The Student Understanding of Science and Scientific Inquiry (SUSSI) instrument was used to measure NOS views across all study participants during both semesters. In particular, data collected from the SUSSI was used to answer researcher questions 1, 3, and 3a. The SUSSI consists of 24 Likert scale items and 6 open response questions. The items are separated into 6 NOS components. Each component consists of 4 Likert scale items and 1 open response question. The six NOS concepts included on the SUSSI are observations and inferences, the tentative nature of scientific theories, scientific laws and theories, the social and cultural influence on science, imagination and creativity in scientific investigations, and science methodology.

The SUSSI was selected for three primary reasons. First, it is more conducive to inferential statistics than alternative instruments. Since there is limited empirical evidence for the Klassen (2009) story framework, it was important to be able to take advantage of the relatively large total sample size (*n*= 183) in the current study through using inferential statistics.

Significant findings from the quantitative data complemented by the qualitative findings of the current study would provide support for conducting more in-depth qualitative studies of the Klassen (2009) framework in the future. Second, using an established instrument was an important consideration for the current study to allow for comparisons with the large amount of previous NOS studies. Most of these studies has used instruments based on the multidimensional NOS framework such as the SUSSI and VNOS (Deng et al., 2011). Finally, as was noted in chapter 2, there are no comparable valid and reliable instruments based on the Argumentative Resource (AR) framework currently available. It could be argued that an instrument developed based on the AR framework would be a better fit for evaluating the effects of the Klassen (2009) framework. However, a new instrument would have needed to have been developed and validated which was outside the scope of the current study. The SUSSI was therefore the best fit for the objectives and research questions out of the currently available instruments.

Additionally, the SUSSI was validated and checked for reliability with the population intended for this study. The initial development of the SUSSI consisted of four phases. First, a pool of 58 items was created based on reform documents (AAAS, 1993; NRC, 1996; McComas & Olson, 1998) and three previous NOS instruments (Aikenhead & Ryan, 1992; Lederman et al., 2002; Chen, 2006). The initial item pool was then subjected to several rounds of pilot tests with pre-service elementary teachers, in-service elementary teachers, science education researchers, and undergraduates. Feedback from the pilot tests was used to revise item wording and develop scoring schemes for the Likert items. The pilot tests resulted in the 24 items included on the final version of the SUSSI. The final version of the SUSSI was then again field tested with a larger sample of pre-service elementary teachers. The results of the second field test were used to further modify the wording of the Likert and open-ended response questions to improve agreement between participant responses to each question type. SUSSI was also validated with international populations of students in China and Turkey in a subsequent study (Liang et al., 2009). Validation on international populations increases the flexibility of the instrument. Reliability of the instrument was measured using Cronbach's alpha, which was 0.69 for the whole instrument.

The SUSSI was selected after consideration of several other instruments. There is a long history of researchers developing instruments for measuring NOS views. Many older instruments from the 1950s through the 1980s are no longer appropriate for use. These older instruments tend to focus on aspects of science outside of current conceptions of NOS such as participants' feelings and attitudes toward science. This is particularly problematic in that several of the instruments only provide a unitary score to represent NOS views. Examples of these older instruments include the Science Attitude Questionnaire (Wilson, 1954), Facts About Science Test (FAST) (Stice, 1958), Processes of Science Test (BSCS, 1962), Science Support Scale (Schwirian, 1968), Science Attitude Inventory (SAI) (Moore and Sutman, 1970), and the Science Inventory (SI) (Hungerford and Walding, 1974).

As discussed in chapter 2, there are three relatively recent instruments that have been rigorously validated with secondary or college students, all of which use a conceptualization of NOS consistent with the multidimensional framework interpretation of NOS (Deng et al., 2011). These instruments are the Views on Science-Technology-Society (VOSTS) (Aikenhead & Ryan, 1992), Views of Nature of Science Questionnaire (VNOS) (Lederman et al., 2002), and the Student Understanding of Science and Scientific Inquiry (SUSSI) (Liang et al., 2008). However, there are concerns with the VOSTS and VNOS for larger scale quantitative studies of NOS views. The VOSTS has redundant and ambiguous items that in some cases may not discern the nuance in participants NOS views (Liang et al. 2008). The VNOS requires participants to complete 10 open response items over 45-60 minutes which can be problematic for participants with limited writing skills and knowledge of NOS. It also makes studies with large amounts of participants impractical. As a result, the SUSSI was deemed to be the best fit for this study.

Quantitative data was also collected related to students' understanding of Mendelian genetics using a two-tier instrument developed by Tsui and Treagust (2010). The genetics two-tier instrument data is important for answering research questions 2. The instrument consists of 12 items related to monohybrid crosses, inheritance patterns, and family pedigrees. These content areas align with the previously mentioned content covered in the genetics unit. Each item consists of two parts. The first part asks participants to respond to a multiple-choice question regarding a genetics concept. The second part consists of a multiple-choice question asking the participant to explain their reasoning for the first part of the item. There is a pre- and post-instruction version of the instrument. The two versions are the same with the exception that different examples and numbers are used in the items. The Tsui and Treagust two-tier instrument was selected for several reasons. First, it is relatively easy to administer during class time and provides data that is conducive to inferential statistics. According to David Treagust (1988), one of the original users of two tier instruments in science education, participants should only receive credit for a correct response if they answer both parts of an item correctly. If the participants only answer the content question or the explanation question correctly, then the item is considered incorrect. The mean overall scores can then be compared pre- to post-instruction and or across the intervention and control treatments (Tsui & Treagust, 2010).

Additionally, the Tsui and Treagust (2010) two-tier instrument is the only recently developed, valid, and reliable instrument that focuses specifically on concepts related to Mendelian genetics[1]. The validation of the instrument was done in Australia. The initial pool of items was developed based on Australian and international genetics literature as well as Australian textbooks, curriculum documents, and college entrance exams. The items were intended to measure student genetics reasoning related specifically to Mendelian and molecular genetics. The initial two-tier items were reviewed by 2 university level lecturers and 2 science teachers in Australia. These experts helped to establish content validity and their feedback led to several rounds of revision. After expert review was completed, the items were piloted with a preservice teacher and a 10th grade student. Feedback from both of these individuals was used to further refine the items. The resulting items were then used at 3 secondary schools successively. Between each administration of the items at the secondary schools, the items were refined based on feedback from participating teachers and students at the schools. The results from the pilot testing of the instrument at the 3 Australian secondary schools found that the instrument was discriminating between and among 10th and 12th grade students. The instrument was also found to be reliable based on Cronbach's alpha (pre-test $\alpha = 0.75$; post-test $\alpha = 0.64$). While the instrument is not validated with college students, it is likely to be valid and reliable with these students as well since introductory college students often have similar genetics misconceptions to high school students (McElhinny et al., 2014).

Qualitative data. The primary means of qualitative data collection for this study was semi-structured interviews with a sub-sample of participants from both the intervention and control groups. 15% of all participants were interviewed from the control and intervention groups. The 15% threshold was chosen because previous studies in NOS research and qualitative

[1] All information in this paragraph related to the validity and reliability of the two tier instrument is from Tsui and Treagust (2010) unless otherwise noted.

research more generally, have shown interviewing 15-20% of a sample is sufficient for reaching saturation of new information (Lederman et al., 2002; Guest et al., 2006). The interviews were used in answering research questions 3-4. The interview questions covered two primary areas including impressions of the class format and participant responses on both the SUSSI and the two-tier genetics instrument. The full interview protocol is included in Appendix C. Interviews were conducted in Wood Hall on the campus of Western Michigan University and lasted around 20-30 minutes each. The interviews were audio recorded. All interviews were conducted during the last two weeks of the semester for both the intervention and control groups. Conducting the interviews at the end of the semester for both groups allowed for the intervention and all data collection instruments to be administered in the intervention group and maintained symmetry with the control group.

Instructional Fidelity

All class sessions for both the control and intervention groups were taught by the same instructor. The course instructor is an experienced biology instructor with extensive experience conducting science education research. Additionally, the primary researcher observed all class periods in which the historical story intervention was administered. The observations were intended to ensure that the intervention story was fully administered in the intervention group.

Data Analysis

This study followed a quasi-experimental design with a nonequivalent control group. Participants were not randomly assigned to the control or intervention group. However, they were not able to choose a group and were unaware of whether they were participating in the control or the intervention group. Both the SUSSI and the two-tier genetics instruments were used in a pre-post fashion at the beginning and end of the semester for both groups. For the qualitative portion of the study, semi-structured interviews were conducted with a sub-sample of control and intervention group participants after instruction at the end of the semester. Details regarding the analysis procedures for both portions of the study are included below.

Table 7

SUSSI Likert Items Comprising NOS Components

NOS Component	Items (Scoring)
Observations and Inferences	**1A** Scientists observations of the same event may be different because the scientists' prior knowledge may affect their observations. (+) **1B** Scientists observations of the same event will be the same because scientists are objective. (-) **1C** Scientists observations of the same event will be the same because observations are facts. (-) **1D** Scientists may make different interpretations based on the same observations. (+)
Tentativeness	**2A** Scientific theories are subject to on-going testing and revision. (+) **2B** Scientific theories may be completely replaced by new theories in light of new evidence. (+) **2C** Scientific theories may be changed because scientists reinterpret existing observations. (+) **2D** Scientific theories based on accurate experimentation will not be changed. (-)
Scientific Laws and Theories	**3A** Scientific theories exist in the natural world and are uncovered through scientific investigations. (-) **3B** Unlike theories, scientific laws are not subject to change. (-) **3C** Scientific laws are theories that have been proven. (-) **3D** Scientific theories explain scientific laws. (+)
Social and Cultural Influence	**4A** Scientific research is not influenced by society and culture because scientists are trained to conduct pure, unbiased studies. (-) **4B** Cultural values and expectations determine what science is conducted and accepted. (+) **4C** Cultural values and expectations determine how science is conducted and accepted. (+) **4D** All cultures conduct scientific research the same way because science is universal and independent of society and culture. (-)
Creativity and Imagination	**5A** Scientists use their imagination and creativity when they collect data. (+) **5B** Scientists use their imagination and creativity when they analyze and interpret data. (+) **5C** Scientists do **not** use their imagination and creativity because these conflict with their logical reasoning. (-) **5D** Scientists do **not** use their imagination and creativity because these can interfere with objectivity. (-)
Scientific Methodology	**6A** Scientists use different types of methods to conduct scientific investigations. (+) **6B** Scientists follow the same step-by-step scientific method. (-) **6C** When scientists use the scientific method correctly, their results are true and accurate. (-) **6D** Experiments are not the only means used in the development of scientific knowledge. (+)

Note. Adapted from Liang et al. (2008). Items marked (+) are scored positively from strongly disagree (1) to strongly agree (5). Items marked (-) are scored negatively from strongly agree (1) to strongly disagree (5).

Quantitative data. All participant responses to the SUSSI Likert items were scored from 1-5 from the least informed view to the most informed view. A scoring scheme developed by the original SUSSI authors was used to score the Likert items (see Table 7 for the scoring scheme) (Liang et al., 2008). A composite score for each of the six SUSSI components was calculated for each participant. The SUSSI composite scores are a sum of the scores for the 4 Likert items for the component. Mean composite scores were calculated for all participants. Following the approach of researchers who have used the SUSSI for previous studies, multivariate analysis of variance (MANOVA) was used to compare pre- and post-instruction mean composite scores for the control and intervention group (Miller et al., 2010; Clough et al., 2010; Park et al., 2014). Post-hoc tests were used to identify particular NOS components that were significantly different pre- to post-instruction. The magnitude of any differences observed between groups was calculated using partial Eta-squared. This measure of effect size was selected because it is the most commonly used measure for education research (Richardson, 2011).

For the genetics two-tier instrument data, scores were calculated for each participant using a scoring key provided by one of the original authors (C. Tsui, personal communication, August 31, 2014). Participants received a point for each item that they correctly answered both the first and second tier questions. No points were assigned to items in which participants answered only the first or the second tier question correctly. Paired t-tests were used to compare overall pre- and post-test scores within the treatment groups. The paired t-tests determined whether there were significant differences in students' genetics understanding before and after instruction. Similarly, independent t-tests were used to compare overall scores across the treatment groups. This determined whether there were any statistically significant differences between the intervention and control groups. McNemar's test was used to compare individual items on the pre- and post-two-tier genetics instrument. In addition, effect sizes were calculated for the paired t-test and independent t-test results to determine the magnitude of any differences between groups. The results from the Tsui and Treagust two-tier genetics instrument were used to answer research question 2.

Qualitative data. SUSSI open response items were scored using a scheme developed by the original authors of the instrument (Liang et al., 2008; Miller et al., 2010). Each response was rated from naïve to informed on a 3-point scale. Consistent with prior research on intercoder reliability, two researchers coded 10% of all the open responses independently. Once an intercoder agreement of 75% was established, the primary researcher then coded the remaining open responses (Campbell et al., 2013). Frequency counts, means, and standard deviations were calculated for the open response questions. The open response data was used for answering research question 1.

All interviews were audio recorded and transcribed by the primary researcher and a research assistant. The interview data was then coded typologically to determine whether and how students' NOS and genetics understanding changed as a result of instruction in the control and intervention groups (Marshall & Rossman, 2011). HyperRESEARCH qualitative analysis software was used for all coding (Version 3.7.2, Researchware, Inc., 2015). The software allows for digital organization and storage of the qualitative data and results. It also allows for easier development of reports of code usage which can be useful for developing themes. The coding process began with the primary researcher developing an initial coding scheme (codebook) through multiple readings of the interview transcripts[2]. The initial codebook included tentative codes, definitions, and any rules for code application. An iterative process was then used to refine the codebook. The primary researcher shared the initial coding scheme with a secondary researcher who provided feedback on the coding scheme. The coding scheme was then revised based on the feedback given by the secondary researcher. Several transcripts were then chosen randomly and coded by the primary and secondary researcher independently. The researchers subsequently met to discuss their coded transcripts. All discrepancies or disagreements served as opportunities to revise and refine codes, definitions, and rules. This iterative codebook refinement process of independent coding and discussion of discrepancies between researchers continued until the primary researcher was satisfied that no new information was emerging.

Data Integration

The results for the quantitative and qualitative portions of the study were integrated after the analysis was completed for both separately (Leech & Onwuegbuzie, 2009). The combined datasets were used to answer research questions 3 and 3a. In addition, they were used as a check

[2] The codebook development process outlined in this paragraph is based on processes described in Campbell et al. (2013) and Weston et al., (2001).

on the validity of the SUSSI instrument. Special attention was paid to the agreement between NOS understandings revealed by the SUSSI and those revealed by the interviews.

Validity and Reliability

Since the current study utilized mixed methods, validity and reliability needed to be evaluated for both the quantitative and qualitative data. The procedures used for checking validity and reliability for both portions of the study are discussed separately below.

Quantitative. Validity and reliability were considered for both quantitative instruments and the overall design of the quantitative portion of the study. In order to ensure the quality of a study, two general types of validity that need to be evaluated are internal and external validity. Several specific types of validity that fall under these two general categories are considered below. As discussed previously, the validity and reliability of both quantitative instruments was established previously by the original authors of the instrument (Liang et al., 2008; Tsui and Treagust, 2010). This being said, reliability and validity was established for both instruments for the participants of the current study. Reliability will be measured for both instruments using Cronbach's alpha, a well known measure of instrument reliability. It is important to reestablish reliability for each sample to ensure that data collection instruments are providing consistent measurements.

Internal validity refers to the extent that differences in the dependent variable measured in a study are the result of the independent variable. Onwuegbuzie and McLean (2003) noted 43 potential threats to internal validity at the data collection and analysis stages of a given study as part of his Quantitative Legitimation Model. Seven threats that are most relevant to the current study are discussed here. First, there are 3 threats to internal validity related to participants' response to participating in a research study. Testing validity refers to the potential of study participant scores being improved by the presence of a pre-test. By simply having taken the test once already, participants may have an advantage on the post-test. Testing validity was mitigated in the current study because both the control group and the intervention group received pre-tests. Additionally, there was considerable time between the pre- and post-tests since they were given at the beginning and end of the semester respectively.

A related threat is evaluation anxiety which is error introduced by participants experiencing test anxiety. Evaluation anxiety was minimized in the current study since both quantitative measures were given in a non-evaluative manner without strict time limits. Participants were informed that they would receive credit for simply completing the pre- and post-tests and not based on whether their responses were correct.

The behavior and performance of participants can also be affected simply by their knowledge of participating in research. This threat to internal validity can be referred to as reactive arrangement. Examples of reactive arrangements are the Hawthorne effect and placebo effect where participants perform better because they believe they are receiving special treatment. The opposite effect, referred to as the John Henry effect, can also be problematic if participants that are aware that they are in the control group gain extra motivation to perform up to the level of the intervention group. Reactive arrangement threats were mitigated in the current study because participants were not aware of whether they were assigned to the intervention or control group. In addition, the treatment groups were conducted in two different semesters making communication between groups unlikely.

In regard to the instruction received by participants, implementation bias can introduce problems if participants do not receive the same instruction across instructors. Implementation bias becomes more likely as the number of instructors involved in a study increases. Implementation bias was minimized in the current study as only one instructor delivered instruction to both groups.

For data analysis, differential selection of participants occurs when the control and intervention groups are not equivalent and therefore not suitable for comparison. As stated previously, demographic data was collected in the current study to ensure that the comparison groups were sufficiently similar. Another threat during the data analysis stage is researcher bias. Prior knowledge of participants can bias analysis, particularly during analysis of open response questions. However, the primary researcher in the current study was not the participants' instructor and therefore had limited prior knowledge of the participants during analysis. Additionally, during analysis of the SUSSI open response questions researchers were kept unaware of whether responses were from the pre- or post-SUSSI.

External validity refers to the generalizability of results to groups or populations outside of the sample population. Many of the threats discussed above also apply to external validity.

The size of the sample for the current study, 187 participants, means there may be some generalizability to the population of students taking Biology 1120 at Western Michigan University. However, the sample was a non-random convenience sample meaning that there will be limited generalizability to any outside groups or populations. Addressing the limited generalizability of the current study will require additional replicative studies in the future (Onwuegbuzie & McLean, 2003).

Qualitative. Onwuegbuzie and Leech (2007) provide an analogous model for determining validity in qualitative research. Some qualitative researchers are uncomfortable with using the term validity because they believe it assumes there is a single reality. As a result, Onwuegbuzie and Leech's (2007) Qualitative Legitimation Model uses the term internal credibility to refer to issues of validity to be sensitive to this concern and reflect the differences between qualitative and quantitative research. Internal credibility is the "…truth value, applicability, consistency, neutrality, dependability, and/or credibility of interpretation and conclusions within the underlying setting or group" (Onwuegbuzie & Leech, 2007, p. 235). There are 14 threats to internal credibility noted by the model authors. The 9 threats most relevant to the current study are discussed here.

There are 4 threats related to providing an accurate account of participant experiences that is supported by the data. Voluptuous legitimation refers to the degree to which research interpretations of the data exceed what is actually shown by the data. Descriptive validity is the factual accuracy of the account provided by the researcher. Structural corroboration is the extent to which multiple data types are used to argue for or against interpretations. Meanwhile, theoretical validity is the extent that the researcher's theoretical explanations fit the data.

There are also 3 threats that stem from not fully considering alternative explanations for the results. Ironic legitimation is the idea that there are multiple realities and that quality research is able to represent "co-existing opposites." Confirmation bias happens when a researcher's interpretations and conclusions follow *a priori* hypotheses too closely. It can become problematic if there are viable alternative explanations that are not fully considered by the researcher. Causal error occurs when researchers give causal explanations for phenomena without verifying these explanations with the data.

Finally, there are 2 threats that involve the personal biases of the researcher and the participants' involvement in the research respectively. Researcher bias occurs when the researcher has *a priori* assumptions that they are unable to bracket. These biases can be transferred to participants or influence the researcher during data collection and analysis. Reactivity is changes in participant performance based on their knowledge that they are participating in research.

Several strategies can be used to mitigate the above described threats to credibility. There are 4 such strategies that were used in the current study including data triangulation, weighting data sources, using a control group, and investigating outliers. Data triangulation reduces the likelihood of identifying false associations in the data and of biases being introduced by any particular method used in the study. The current study used method triangulation (quantitative and qualitative), data triangulation (multiple sources), and investigator triangulation (multiple researchers were involved in data analysis). Weighting data involves assigning greater weight to higher quality data sources and increases confidence in the inferences and conclusions made based on the data. Using a control group can help with mitigating the threat to creditability from reactivity. The current study used a control and intervention group in which both groups were aware that they were participating in research. As a result, it is unlikely that any differences observed were the result of reactivity. Finally, investigating outliers involves fully exploring findings that do not follow previous patterns in the data. Paying close attention to outliers in the data can help with mitigating ironic legitimation, confirmation bias, and causal error.

Limitations

The current study has the following limitations. First, the choice of forced choice instruments for evaluating students' NOS and genetics understanding may have not allowed students to fully or accurately describe their understandings. However, this limitation was at least somewhat mitigated by the use of semi-structured interviews, which allowed a sub-group of participants to expand upon and or explain their responses to the SUSSI and two-tier genetics instruments. Additionally, the sample was one of convenience and not randomized. Both of these limitations do not allow for generalizing the results to external groups or populations. The participants in this study did, however, provide insight into the influence of the intervention and the particular portions of the intervention that may have been most helpful for them. They may have also provided insights into the population of students taking Biology 1120 at Western Michigan University based on the sample size. Another potential limitation of the study is the length of the intervention itself. It could be argued that two class periods are not enough time for students to achieve adequate NOS understandings. However, statistically significant differences were observed on the SUSSI from pre- to post-instruction for the intervention group. Participants also pointed towards historical stories in the course being a large influence in changing their NOS views. Chapters 4 and 5 will discuss the findings of the study in detail. Additionally, previous studies have shown that shorter interventions can have positive effects on NOS views (Kim and Irving 2010; Rudge et al. 2014; Williams & Rudge, 2016).

CHAPTER IV

RESULTS AND DISCUSSION

In the previous chapter, the methods used to conduct the current study were described. This chapter presents the results of the study and discusses the meaning of the results related to each of the research questions noted previously. Additionally, analysis done to establish validity and reliability of the quantitative instruments for the study sample is included. The chapter is organized by individual research questions. All quantitative analysis was done using SPSS software (version 24).

Research Question 1

What differences in NOS understandings are revealed from pre- to post-instruction based on participants' SUSSI Likert and open response scores in both the traditional and historical story groups?

Quantitative measures of SUSSI validity and reliability. As noted in chapter 3, validity and reliability were established for the current study sample. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to establish construct validity for the six NOS aspects on the SUSSI using a combined data set with all data collected in both treatment groups (n= 454). The results of the EFA showed relatively strong loadings for all of the NOS aspects except for the sixth, Methodology of Scientific Investigations (Table 8).

Table 8

SUSSI Exploratory Factor Analysis Item Loadings

SUSSI Item Number	Observations and Inferences	Change of Theories	Laws vs. Theories	Social and Cultural Influence	Creativity and Imagination	Scientific Methodology
1A	.394					
1B	.662					
1C	.509					
1D	.180					
2A		.655				
2B		.599				
2C		.677				
2D		.350				
3A			.201			
3B			.279			
3C			.417			
3D			-.259			
4A				.394		
4B				.740		
4C				.630		
4D				.415		
5A					.701	
5B					.779	
5C					.904	
5D					.884	
6A						.607
6B						.458
6C			.333			
6D				.192		

Two of the items that were expected to load on the factor for Methodology of Scientific Investigations, 6C and 6D, loaded on other factors. Given that the majority of the remaining items strongly loaded onto their expected factors, the results of the EFA supported further analysis of the construct validity of the six factors using CFA.

The CFA was conducted using the AMOS package for SPSS. CFA requires that you define a model prior to running the analysis. The model was defined as having six factors with four items each. The six factors and their four items corresponded with the six SUSSI aspects and their accompanying items from the SUSSI instrument. Goodness of fit for the hypothesized model was evaluated using several measure guidelines including the root-mean square error of approximation (RMSEA; Steiger, 2000), root mean square residual below .09 (RMR; Hu and

Bentler, 1999), chi-squared/*df* below 5.0 (Bollen, 1989), and a comparative fit index (CFI) near 0.90 (Hu and Bentler, 1999; Byrne, 2013). Guidelines for acceptable model fit statistics values for RMSEA vary. Hu and Bentler (1995) suggest an RMSEA of 0.06 as indicative of a good-fitting model. MacCallum et al. (1996) suggest values of 0.01, 0.05, and 0.08 as indicative of excellent, good, and mediocre fit, respectively. CFA model fit statistics are included in Table 9 below. All of the model fit statistics indicate strong model fit and support the six-factor model. The data from this study sample support underlying constructs for the six NOS aspects included by the original authors.

Table 9

SUSSI Confirmatory Factor Analysis Model Fit Statistics

Model Fit Statistic	Value
RMSEA	.05
RMR	.055
chi-squared/*df*	2.13
CFI	.879

Internal reliability of the SUSSI was established for the current study using Cronbach's alpha. Overall reliability for each administration of the SUSSI was within acceptable ranges for social research (Hatcher and Stepanski, 1994). These overall reliability values are also close to the numbers the original SUSSI authors reported in the development of SUSSI (Liang et al., 2008). The alpha values for the Laws vs. Theories and Scientific Methodology aspects were both quite low. While this is consistent with the reliability values reported by the original SUSSI authors, any significant findings related to these aspects should be considered with caution. Together the factor analysis and reliability results indicate that overall the SUSSI was a valid and reliable instrument for the current study sample.

Table 10

SUSSI Cronbach's Alpha Values

SUSSI NOS Aspect	Control Pretest	Control Posttest	Intervention Pretest	Intervention Posttest
Overall SUSSI	0.624	0.729	0.749	0.712
(1) Observations and Inferences	0.486	0.524	0.539	0.445
(2) Change of Scientific Theories	0.615	0.687	0.618	0.601
(3) Laws vs. Theories	0.126	0.194	0.277	0.174
(4) Social and Cultural Influence	0.673	0.663	0.548	0.641
(5) Creativity and Imagination	0.876	0.919	0.875	0.889
(6) Scientific Methodology	0.433	0.255	0.346	0

SUSSI results. Paired SUSSI data were used to make comparisons between pre- and post-instruction NOS views for both the control (*n*= 91) and intervention (*n*= 92) groups. Pre- and Post-instruction composite scores for each of the six SUSSI NOS aspects were compared using multiple analysis of variance (MANOVA). The composite scores were combined scores for the four Likert items under each SUSSI NOS aspect. Figure 2 and Figure 3 below show descriptive statistics and results of significance testing for the control and intervention groups respectively. For both the control and intervention groups, the MANOVA results suggested that there were statistically significant differences between the pre- and post-instruction scores for one or more of the SUSSI NOS aspects. The control group had Wilk's Lambda value = .758, $F(6,85)= 4.52$, $p= .001$, and partial $eta^2 = .242$. The intervention group had Wilk's Lambda value = .704, $F(6, 86)= 6.04$, $p<.001$, and partial $eta^2 = .296$. Effect size for both groups was measured by partial eta^2. The partial eta^2 values for both the control and the intervention groups indicate a large effect (Cohen, 1969; Richardson, 2011). This suggests that there was a significant change in SUSSI composite scores from pre- to post-instruction for both groups that was influenced extensively by testing occasion. There were no significant differences detected between control group pre-scores and experimental group pre-scores or control group post-scores and experimental group post-scores.

Post-hoc tests using Sidak's Test for multiple comparisons showed statistically significant decreases for the control group in mean composite SUSSI scores for the Observations and Inferences (p= .004) and Social and Cultural Influence aspects (p=.039). Mean scores for Observations and Inferences decreased from 15.82 (SD= 2.55) on the pre-SUSSI to 14.91 (SD=

2.65) on the post-SUSSI. The Social and Cultural Influences mean scores decreased from 14.55 (SD= 3.12) on the pre-SUSSI to 13.81 (SD= 3.08) on the post-SUSSI. For the intervention group, a statistically significant increase in SUSSI mean composite score was detected for the Imagination and Creativity aspect (p< .001). Mean scores for the Imagination and Creativity aspect increased from 11.09 (SD=4.00) on the pre-SUSSI to 12.45 (SD= 3.81) on the post-SUSSI. These results suggest that something about the experience of the control group participants may have clouded their understanding of observations and inferences in science and the role of culture and society in science. In contrast, the intervention group participants were able to significantly improve their understanding of the role of imagination and creativity in science from pre- to post-instruction. Interview data indicated that the instruction received by participants and their experiences in the BIOS 1120 course played a role in influencing these changes. These results will be discussed further later in this chapter.

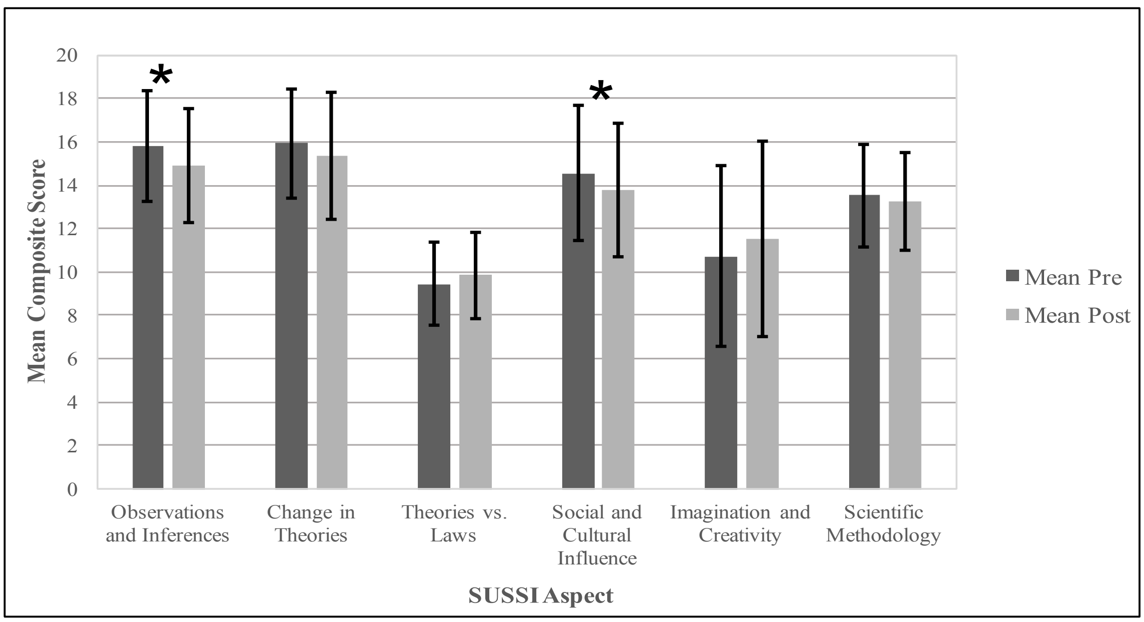


Figure 2. SUSSI mean composite scores for control group. Error bars represent standard deviation. *indicates significant decrease from pre- to post-SUSSI (p<.05).

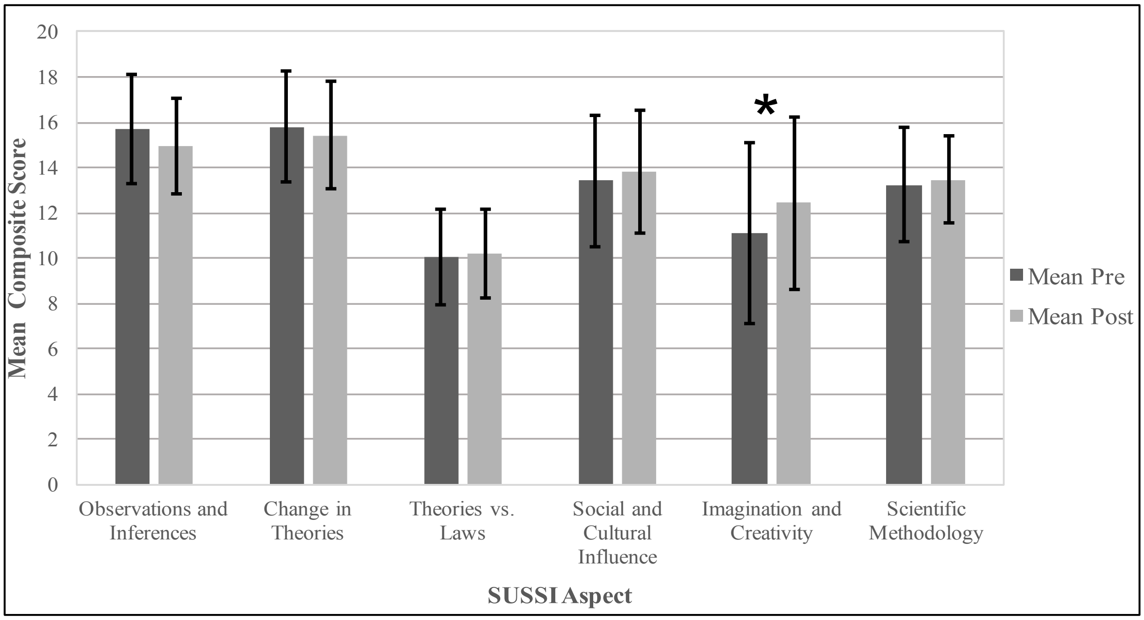


Figure 3. SUSSI mean composite scores for the intervention group. Error bars represent standard deviation. *significant increase from pre- to post-instruction score ($p<.001$).

The SUSSI open response item scores were also examined to determine if they revealed any additional differences in NOS views from pre- to post-instruction for the control and intervention groups. General trends in the SUSSI open response scores aligned with scores on the SUSSI Likert items. Those NOS aspects that had the highest mean composite scores for the Likert items also had the highest proportion of open responses that received a score of 2 (transitional view) or 3 (informed view). The NOS aspects with the lowest mean composite score, Theories vs. Laws, had the highest proportion of open responses receiving a score of 1 (naïve view). SUSSI open response pre- and post-scores were then compared using McNemar's Test to determine if there were any significant differences for any of the six NOS aspects. No significant differences were detected with the exception of the Change in Theories aspect which saw a significant decrease in scores for the intervention group. The lack of significant difference in the open response items is likely at least in part due to the scoring scheme created by the SUSSI authors. Several of the scoring criteria are quite arbitrary which leads to a high number of responses being coded as transitional as a kind of a catch all default.

Summary of SUSSI Likert results. The results of the validity and reliability tests conducted for the SUSSI instrument for the current study sample showed that overall the instrument was valid and reliable for the study sample. Comparisons between pre- and post-

SUSSI composite scores completed using MANOVA indicated significant differences from pre- to post-instruction in NOS understanding for both the control and intervention groups. The control group showed significant decreases in mean composite scores for the Observations and Inferences and Cultural and Societal Influence aspects. The intervention group had a significant increase in mean composite scores for the Imagination and Creativity aspect. SUSSI open response item scores were generally aligned with Likert items. However, there was a lack of alignment between significance test results for the SUSSI Likert composite scores and open response scores. The primary difference between the treatment groups was the introduction of historical stories in the intervention group. These findings indicate that the historical stories positively influenced the NOS views of intervention group participants related to imagination and creativity in science.

Research Question 2

Q2. What differences in Mendelian genetics understandings are revealed from pre- to post-instruction based on participants' two-tier survey scores in both the traditional and historical story groups?

Quantitative measures of two-tier genetics instrument validity and reliability. Content and face validity of the two-tier genetics instrument used in this study were established through the extensive process of the original authors described in chapter 3 (Tsui & Treagust, 2010). Several standard test analysis measures were used to establish item validity and reliability for the current study sample. These measures included facility index, discrimination index, and Cronbach's alpha. All three of these test analysis measures have been commonly reported for similar two-tier instruments in the science education literature (Tan et al., 2002; Odom & Barrow, 1995). The facility index refers to the difficulty of an item and is calculated by adding the number of high performing and low performing participants who answered an item correctly and dividing by the total number of high and low performing participants. High performing participants are those that scored in the top 27% on the overall test and low preforming participants are those in the bottom 27%. Discrimination index is a related measure of how well an item differentiates between high and low performers. It is calculated by subtracting the number of low performers that answer an item correctly from the number of higher performers that answer the item correctly and dividing by 27% of the total sample size. Both the facility index and discrimination index produce values between 0 and 1 with higher numbers signifying easier and more discriminant items respectively.

Tables 11 and 12 below contains a breakdown of the facility index, discrimination index, and Cronbach's alpha measures for all administrations of the two-tier genetics instrument for the current study. The control group had an average facility index on the pre-test of .37 and on the post-test of .43. It had an average discrimination index of .40 on the pre-test and .35 on the post-test. The intervention group had an average facility index of .34 on the pre-test on .41 on post-test. While the average discrimination index was .40 on the pre-test and .43 on the post-test.

Together these values indicate that the instrument was difficult for participants on all testing occasions and able to discriminate between those participants with higher and lower genetics content understanding. The genetics two-tier instrument had items ranging from easy where over half of participants answered correctly to items that were quite difficult where only 10% to 30% of participants answered correctly. Researchers developing similar types of two-tier instruments have set .20 as a benchmark for discrimination index (e.g. Odom and Barrow, 1995). Items with discrimination indexes over .20 are considered adequately able to differentiate between high and low performers. The majority of items in current study had discrimination indexes well above .20 for all testing occasions.

Table 11

Control Group Two-Tier Genetics Instrument Test Analysis Measures and Reliability (n=77)

Measures	Pretest	Posttest
Number of items with facility index (x)		
.7 < x < .9	1	3
.5 < x < .7	3	1
.3 < x < .5	2	3
.1 < x < .3	6	5
Number of items with discrimination index (y)		
.5 < y < .7	5	1
.3 < y < .5	3	6
.1 < y < .3	4	5
Cronbach's alpha	.590	.489

Table 12

Intervention Group Two-Tier Genetics Instrument Test Analysis Measures and Reliability (n=90)

Measures	Pretest	Posttest
Number of items with facility index (x)		
.7 < x < .9	1	2
.5 < x < .7	2	2
.3 < x < .5	4	4
.1 < x < .3	3	3
x < .1	1	1
Number of items with discrimination index (y)		
y > .7	1	1
.5 < y < .7	4	5
.3 < y < .5	3	3
.1 < y < .3	2	2
y < .1	2	1
Cronbach's alpha	.591	.619

Two-tier genetics instrument results. In order to answer the research question, comparisons were made between pre- post-instruction overall scores for both the control group and experimental group. Participants were given one point for each two-tier item they answered correctly for a maximum possible total score of 12. First, pre- and post-overall scores were compared for both groups using paired t-tests. Significant differences were detected for both groups from pre- to post-instruction. For the control group, mean overall scores significantly increased from 4.35 on the pre-test to 5.08 on the post-test (*n*= 77*,* p=.006). Cohen's d for the control group was .384 indicating a small to medium effect. The intervention group also had a significant increase in mean overall scores from 3.97 on the pre-test to 4.94 on the post-test (*n*=90, p<.001). Cohen's d was .47 for the intervention group indicating a medium effect. Independent t-tests were also employed to compare pre- and post-instruction overall scores between the control and experimental groups. No significant differences were detected between the pre-scores of control and intervention group or the post-scores.

The significant findings from the comparison of overall scores for the two-tier genetics instrument called for further analysis to determine whether any individual items were significantly different from pre- to post-instruction for both groups. McNemar's Test was used to analyze individual items. Table 13 below includes a breakdown of the number of participants with correct responses for each item on the pre- and post-test.

Table 13

Number of Correct Responses on Genetics Two-Tier Instrument Items

	Control (n= 77)		Intervention (n= 90)	
Item	Pretest	Posttest	Pretest	Posttest
1	61	68	70	74
2	56	62	61	64
3	26	30	27	23
4	10	12	20	15
5	9	15	20	23
6	41	48	29	57**
7	14	25*	22	28
8	56	57	52	67*
9	7	16	5	8*
10	15	12	18	38**
11	16	8	10	6
12	28	42*	24	32

* indicates significant difference using McNemar's Test at $p < .05$
** indicates significant difference using McNemar's Test at $p < .001$

The McNemar's test results indicated that there were two items for the control group that had statistically significant improvements in the number of participants answering correctly. For the intervention group, there were four items that showed statistically significant improvements. The items that showed improvements for both groups are included in tables 14 and 15 below.

The control group had statistically significant improvements in two items related to the definition of a gene and basic gene function. In contrast, the four items intervention group participants significantly improved on were all related to Mendelian genetics. Since the genetics instruction delivered to both groups was held consistent with the exception of the introduction of the Mendel historical story in the intervention group, there are two interesting conclusions that can be drawn. First, the introduction of the historical story seems to have allowed for greater gains in genetics understanding. Participants in the intervention group showed significant improvement on more items. In addition, the differences from pre- to post-instruction on items 6 and 10 were greater than anything observed in the control group as evidenced by the p-values being less than .001 for these two items. Second, the introduction of the historical story may have changed the focus of intervention group participants during the genetics unit. Intervention group participants seem to have had a greater focus on concepts related to Mendelian genetics than the control group. Potential reasons for these differences between the two groups from the interviews will be discussed in subsequent sections of this chapter.

Table 14

Two-Tier Genetics Instrument Items with Statistically Significant Improvement in the Control Group

Item #	
7	Which one of the following statements best describes a gene? [ ] 1. The smallest unit of structure in a chromosome. [x] 2. A sequence of instructions that codes for a protein. [ ] 3. A segment in a DNA molecule. [ ] 4. Don't know.
12	PRE-12: The following shows part of a human gene on a DNA molecule: <p>......TTATGAGCCGTAGCATAGGGCTAGTAGTAGCA......</P> What do you know about the "letters" and the order of their occurrence in the gene? (Choose the best option.) [ ] 1. The kind of "letters" tells you the information about the kind of protein that the gene can make. [ ] 2. The order of occurrence of the "letters" determines the information about the protein the gene can make. [x] 3. The kind of "letters" and the order in which they occur determine the information about the protein the gene can make. [ ] 4. Don't know.

Note. All items are from the pre-test version of the two-tier genetics instrument developed by Tsui & Treagust (2010). Only the first tier is included for each item.

Table 15

Two-Tier Genetics Instrument Items with Statistically Significant Improvement in the Intervention Group

Item #

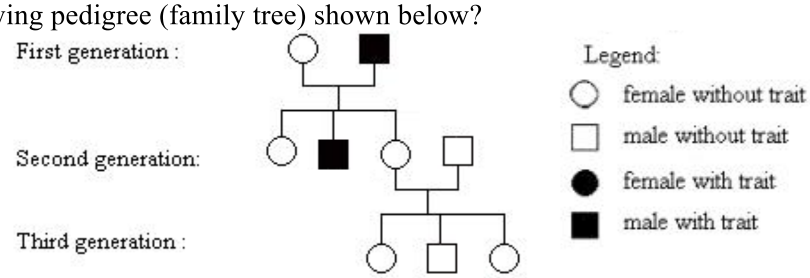
6 In dogs, the gene allele (e) for drooping ears is recessive to E for erect ears. A male dog with genotype Ee was mated to a female dog with genotype ee and gave birth to a litter of 10 puppies. What is the expected proportion of drooping-eared puppies in the litter?

[ ] 1. A quarter
[x] 2. Half
[ ] 3. All
[ ] 4. Don't know

8 In garden pea, white flowers is recessive to purple flower. Suppose we use W for the dominant gene (allele) and w for the recessive gene (allele), what is the genotype of a plant with white flowers?

[ ] 1. Ww
[ ] 2. Ww or ww
[x] 3. ww
[ ] 4. Don't know.

9 Which of the following best describes a common genetic disease using the information in the following pedigree (family tree) shown below?



[ ] 1. Recessive.
[ ] 2. Dominant.
[X] 3. Cannot tell.
[ ] 4. Don't know.

10 In one of Mendel's breeding experiments during the years 1886 and 1887, he cross-fertilized garden peas with round seeds with those with angular seeds (i.e., round x angular). All the offspring peas were round in the first generation (F1). We now know that the F1 peas were hybrids. Mendel crossed two F1 plants and got 732 peas in the second generation (F2). Which of the following might be the F2 phenotypes observed by Mendel?

[ ] 1. 365 round peas and 367 angular peas.
[ ] 2. 185 round peas and 547 angular peas
[X] 3. 547 round peas and 185 angular peas
[ ] 4. Don't know.

Note. All items are from the two-tier genetics instrument developed by Tsui & Treagust (2010)

Summary of two-tier genetics instrument results. Data collected from the current study sample showed that overall the genetics two-tier instrument was a valid and reliable measure of genetics content. Statistical analysis using paired t-tests and McNemar's test revealed that both groups made statistically significant improvements in their understanding of genetics content. The intervention group made more significant improvements in their understanding than the control group. The primary difference between the genetics instruction the groups received was the Mendel historical story. Taken together, there is evidence that the historical stories played a role in allowing the intervention group participants to make greater gains in their genetics content understanding.

Research Question 3

Q3. What types of explanations do participants use for changes in their SUSSI Likert responses from pre- to post-instruction, as revealed by the interviews, in both the traditional and historical story groups?

Explanations for changes in SUSSI responses. A primary focus of the semi-structured interviews conducted with participants was their responses on the pre- and post-SUSSI. Participants were provided with both of their completed surveys during the interview and asked to explain their answers on the SUSSI Likert items. Special attention was paid to those items where participants changed their response from the pre- to post-SUSSI.

Participants in both the control and the intervention groups gave three general types of explanations for changes in their SUSSI responses. Explanation types included historical, general course, and external influence explanations. Historical explanations refer to when participants said that the reason they changed was because of one or more historical examples from the course. General course explanations consisted of participants mentioning some aspect of the course other than historical examples. External influence explanations were when participants referred to something unrelated to the course that influenced their thinking. Table 16 has frequency counts for each of the explanation types for both treatment groups broken down by SUSSI NOS aspect.

Table 16

Frequency Counts of All Explanations Types for SUSSI Likert Response Changes

	Control (*n*=15)			Intervention (*n*= 12)		
SUSSI NOS Aspect	Historical	General	External	Historical	General	External
(1) Observations and Inferences	4	3	3	2	2	—
(2) Change of Scientific Theories	2	—	—	1	2	—
(3) Laws vs. Theories	—	—	—	—	3	1
(4) Social and Cultural Influence	2	2	2	2	2	—
(5) Creativity and Imagination	2	3	1	4	2	1
(6) Scientific Methodology	1	2	3	1	—	—
Total	11	10	9	10	11	2

SUSSI historical explanation results. Six of the fifteen students interviewed in the control group made historical references when explaining changes they made in their SUSSI Likert responses. In the intervention group, seven of the twelve interviewees made reference to historical examples in their explanations. Participants in both groups mentioned history in their explanations of positive and negative SUSSI changes (Table 17). Positive changes were when students went from less informed scores on the pre-SUSSI to more informed scores on the post-SUSSI or stated that their informed views were reinforced. Negative changes were participants moving from more informed to less informed or their less informed views were reinforced.

Table 17

Frequency of Positive and Negative Historical Explanations

	Control		Intervention	
SUSSI NOS Aspect	Positive	Negative	Positive	Negative
(1) Observations and Inferences	1	3	2	—
(2) Change of Scientific Theories	2	—	—	1
(3) Laws vs. Theories	—	—	—	—
(4) Social and Cultural Influence	—	2	2	—
(5) Creativity and Imagination	1	1	3	—
(6) Scientific Methodology	—	1	1	—
Total	4	7	8	1

For the control group, one participant made a positive change in their Likert scores for observations and inferences while three made negative changes based on historical examples. The participant that made a positive change went from drawing conclusions about scientific observations by comparing scientists to non-scientists to comparing scientists from different disciplines to each other. The participant stated that this shift came from learning about the work of past scientists in the course.

All three participants that made negative changes related to observations and inferences felt that the examples of Darwin and Mendel presented in class showed that observations are facts. F13 said that:

> We were just learning about like with Mendel and genetics, he wasn't a biologist he was um an accountant or a statistician not a typical biology degree and Darwin wasn't a biologist he was I forget what his discipline was but it was not a biologist and then just seeing that over and over and over again the themes are that different people observe the same thing and see the same facts but draw different conclusions from it which is why I changed.

Participant F13 saw multiple scientists from different disciplines making the same observations that led to the same facts. F10 said learning about Darwin’s work influenced their thinking:

> Um well I guess when professor Rudge was like talking about [Darwin] and his experiment with the birds in the Galapagos um he noticed that like they saw natural selection like they get weeded out and so that was his observation and later it became fact so maybe that’s why I changed my answer...

Participant F10 learned that Darwin made observations that lead to the currently accepted theory of natural selection. As a result, they concluded that observations are facts.

Two participants from the intervention group made positive changes in their Likert scores related to observations and inferences that they explained using history examples. Both of these participants said that they gained a better understanding of the idea that scientists can make different observations of the same event because of prior knowledge. For example, S5 said, “I think I had changed it to strongly agree because I thought that-- like with the moth thing, they knew that this didn't work. So because of that their prior knowledge helped shape what they did.” For this participant, the moth mystery phenomena story gave them an example of how prior knowledge affects scientists’ observations.

There were only two changes across both groups for items related to change in scientific theories that participants explained with historical examples. Two participants from the control group made positive gains on their Likert scores that they attributed to their learning about the study of evolution. F7 noted that, “Um with evolution the theory is always changing. The theory changed with the discovery of fossils and the geographic support too.” Both of these participants made reference to changing perspectives about the theory of evolution that they learned from the course, which helped them to understand how theories can change. It is likely that these students learned these insights from their textbook as the evolution chapter begins with a brief overview of how the theory of evolution was developed which includes mention of the influence of the fossil record and geologic evidence on evolutionary thought.

Four changes in the influence of society and culture Likert scores were explained using historical examples across both groups. Both of the changes in the control group were negative. Both participants went from thinking that society and culture influenced scientists' work to thinking that it does not. F14 changed their opinion because they felt that Darwin went against prevailing ideas of his time about the origin and evolution of life to develop his theory of evolution. As a result, it was clear to this participant that society and culture do not influence scientists. The perspective shared by F14 is reflective of the way ideas were presented in the textbook used for the course. The introduction to the chapter on evolution presents the progression of thought on evolution in terms of pre- and post-Darwin. The pre-Darwinian thought presented includes incorrect ideas about evolution including theological explanations. Post-Darwin, scientists now had the correct theory for evolution. This account gives the impression that Darwin rescued everyone from their incorrect thinking which is a clear oversimplification.

The other control group participant, F10, changed from thinking it was not possible for scientists to conduct pure and unbiased studies to thinking that they can. F10 felt like it would be too much pressure to expect scientists to conduct unbiased studies. However, the scientists that F10 learned about in class appeared to have no trouble conducting their studies. F10 said, "Well yeah because with the scientists that we learned about in class um... they seemed to be pretty like solid with what they came out with so then I wasn't sure about it. That pressure put on them was intentional." The change in this participant's thinking may have again been a result of the textbook. Descriptions of historical scientists in the textbook were typically limited to the theories that scientists were most famous for and not the messy process necessary to get to those theories.

Both of the society and culture Likert score changes in the intervention group were positive. Both of these participants linked their changes to ideas that they learned from historical stories in the class. S1 changed to thinking that society and culture influences how science is conducted because of what they learned about Mendel’s training and background in the course. S1 stated:

> Probably the Mendel thing again. Just he went to so many different colleges, he did so many different careers, he worked in so many different cities and countries, and creating all these ideas and experiments when a lot of times he was told he couldn't do certain things in certain cities because it wasn't "right" there.

This participant felt that the story of Mendel’s early life and his education showed how society and culture can influence science. The other intervention group participant, S1, had their informed views of the role of society and culture in science reinforced by the moth mystery phenomena story. S1 said:

> I disagree with that completely. Once again, the moth thing, that was an issue, so that's why they were researching it. They're not just going to go research a white moth for no reason, there was a reason behind it. And that's because what's going on in society…

The moth story gave participant S1 a memorable example of how society can influence science.

Two control group participants provided historical examples for changes in their Likert scores for the imagination and creativity portion of the SUSSI. One was a positive change and the other was negative. The participant that made a positive change in their score, F7, felt that the examples of historical scientists presented in the course showed that past scientists had to be creative because there was no other way to do science. When they were asked why they changed F7 responded:

> But then [through] learning about the course the content and Darwin, Mendel, Watson, and Crick like all of these people I learned that these guys had to get creative had to get like measuring to find solutions to these ideas. Today they’d be like oh yeah this because of this but they knew nothing about that.

F7 made progress in their views of creativity as they went from thinking creativity was not necessary in science to understanding that it has an important role. However, F7 seems to believe that creativity was more important to past scientists who had less knowledge than current scientists. The participant that had a negative change, F10, initially did not think there was creativity and imagination in science. Learning about Mendel in the course made this participant start to rethink their position but ultimately F10 maintained that creativity and imagination is not involved in science.

Two of the three historical explanations related to imagination and creativity in the intervention group were linked to positive changes in Likert scores. The other participant had their positive views reinforced. S2 improved their views of using imagination and creativity to collect data. They attributed this change to the Mendel story:

> I don't know. I just think in the story I remember he tried to become a priest or something like that, and then he also went to school, he went to college. And I just remember all those factors playing into when he came up with his theory.

For participant S2, seeing an example of how prior knowledge can influence scientists work showed them that creativity is used in science. The second intervention group participant that made a positive change, S7, stated that the moth story showed them several examples of scientists reinterpreting observations. S7 felt that reinterpreting scientific findings requires creativity. S8 had their positive views of imagination and creativity in science reinforced by the stories included in the course. They made a comment about how the stories in general provided examples that supported their view that scientists have to use creativity in their work to make meaningful contributions to their field.

Finally, there was one participant in the control group that made a negative change on SUSSI Likert items related to scientific methodology. Participant F4 made a general comment about how learning about various scientists in class made them less certain about whether there are multiple methods that scientists can use in their work. There was also one participant in the intervention group that made a positive change in their Likert responses related scientific methodology. S11 stated that the moth story gave them an example of how experiments are not always necessary to develop scientific knowledge.

SUSSI general course explanation results. Nine of the fifteen participants interviewed in the control group made general comments about the course when explaining why they changed their SUSSI responses. In the intervention group, eight of the twelve students used general course explanations for their changes. Participants in both groups used general course explanations for both positive and negative SUSSI changes (Table 18).

Table 18

Frequency of Positive and Negative General Course Explanations

SUSSI NOS Aspect	Control		Intervention	
	Positive	Negative	Positive	Negative
(1) Observations and Inferences	2	1	—	2
(2) Change of Scientific Theories	—	—	1	—
(3) Laws vs. Theories	—	—	1	3
(4) Social and Cultural Influence	1	1	2	—
(5) Creativity and Imagination	2	2	3	—
(6) Scientific Methodology	—	2	—	—
Total	5	6	7	5

There were three general course explanations used in the control group related to observations and inferences. Of these three, there were two that were associated with a positive SUSSI Likert change. F13 said that they were more certain that theories based on accurate experimentation can be changed because of their experiences in the course working with other students, the lectures, and reading the textbook. F13 also went from thinking that scientific observations of the same phenomena will always be the same because they are objective to thinking that this is not always true. F13 explained this change by saying the course made them think more about objectivity in science. The control group participant that made a negative SUSSI change went from thinking that scientific observations were not facts to thinking that they are facts. F15 said that this change came from doing readings for the class, "…after you read about it you know that scientists have to take like observations and observations don't change everybody sees it the same so that would be the reason."

In the intervention group, there were two negative SUSSI Likert changes related to observations and inferences that were attributed to general course explanations. S1 said that initially they disagreed with the idea that scientists will make the same observations of the same event because they were unsure. After their learning in the course, S1 said they decided they agreed. S8 made the same shift in thinking regarding scientists' objectivity which they ascribed to their first college semester and taking several science classes.

There was only one SUSSI change across both groups that had a general course explanation for the change of scientific theories aspect. An intervention group participant, S9, went from being unsure to thinking that scientific theories can change because of examples they heard about in class. This participant said they learned in the course that new technology can lead to new discoveries, “Because I think after a time, like the advance of technology, it becomes advanced, people have more knowledge. It can be different and some new knowledge can affect it I think.” When asked for examples, S9 mentioned that some of Darwin’s theories are still accepted today and others are not.

For the laws vs. theories aspect, there were four SUSSI Likert changes with general course explanations in the intervention group. Two of the explanations were for positive changes and two were negative. S7 made a positive change going from being certain that laws cannot be changed to being unsure because they said they learned about theories in the course and this made them confused about the definition of scientific laws. The other positive change was S13 who said they went from being uncertain about whether theories explain laws to agreeing because they vaguely remembered an example of a theory becoming law in class. Both of the negative changes were participants going from being uncertain to agreeing that scientific laws cannot be changed. Both of these participants mentioned learning this from assessments given in class.

Two control group participants gave general course explanations for changes they made on items in the social and cultural influence on science aspect. One was positive and the other negative. The participant that made a positive change, F8, went from thinking that culture does not have an influence on what science is conducted to thinking that it does. F8 said that they made this change because they read about different cultures doing science differently in the textbook. The control group participant that made the negative change, F15, went in the opposite direction and went from saying that culture does influence scientists to saying that scientists need to put their culture aside when doing their work. F15 said that they remembered the instructor telling the class about scientists putting their culture aside in response to another student’s question.

There were also two changes related to the influence of culture and society in the intervention group that had general course explanations. Both of the changes were positive changes where the participants moved towards thinking that cultural values influence what science is conducted. One participant, S6, changed their thinking because of what they learned about evolution. After learning about evolution in the course, S6 came to the conclusion that whether a scientist will work on controversial issues like evolution will depend heavily on their personal beliefs. The other intervention group participant changed their thinking about the role of culture in modern science because of the eco-systems unit.

The imagination and creativity aspect had four changes for the control group that had general course explanations. Two of them were positive and two negative. Both of the positive changes were participants who were uncomfortable with the idea of imagination and creativity in science who realized that sometimes it is necessary to take new innovative approaches to solving problems in science. Participant F6 said they made this shift because of something they read in the textbook about data interpretation. The other participant, F11, said that they changed their minds because of examples from the course of more than one scientist observing the same thing and coming up with different results. F11 did not have any specifics about the examples but said they made them wonder why scientists would come to different results. F11 then concluded that the imagination and creativity of individual scientists must be playing a role.

Both of the negative changes for the control group related to imagination and creativity were from the same participant. Participant F1 initially thought that imagination and creativity were involved because they were thinking of historical scientists. However, after taking the course F1 realized that modern scientists now know so much that imagination and creativity is not necessary. When asked about specific parts of the course that illustrated this idea F1 said, "I guess a lot of the cell stuff. I mean, obviously, we know almost everything there is to know about cells, and that really changed my thinking…" It seems that something about the way that the cell unit was taught gave participant F1 the impression that cell biology is not a particularly active area of inquiry and therefore no longer requires much creativity. Participant F1's thinking may have been influenced by the textbook as textbooks often place more emphasis on science products than the process of science.

The intervention group had three positive changes for imagination and creativity. Participant S1 indicated that they thought imagination and creativity played a large role in science all along but answered that they were unsure about it on the pre-SUSSI because they thought that would be the answer the instructor wanted to see. S1 stated, "I just put uncertain just because when I first started the class I was thinking, well, is that what they're looking for is something that's black and white?" However, after learning about natural selection and evolution in the course and presumably becoming more comfortable with the course instructor, S1 indicated on the post-SUSSI that scientists do use creativity and imagination in science. Both of the other participants changed from thinking that imagination and creativity is not used in science because it can interfere with logical reasoning to thinking that it is used. Participants S6 and S13 recalled the instructor going over experiments with the class and saying that you need to be creative in order to design and conduct quality experiments.

Finally, there were only two negative SUSSI Likert changes observed for the scientific methodology aspect. Both changes were from participants in the control group. Participants F4 and F5 both went from saying that the scientists do not have to use the same step-by-step method to answering that they do all have to follow the same general method. When asked why, both participants referenced learning about the scientific method in the class. Participant F5 specifically mentioned learning about the scientific method in unit one of the textbook.

SUSSI external influence explanation results. Six of the fifteen control group interviewees gave SUSSI change explanations that were related to something external to the course. The majority of the control group external influence explanations were for positive changes (Table 19). Only three of the twelve interviewees in the intervention group gave externally influenced explanations.

Only the control group participants gave externally influenced explanations for observations and inferences. Participant F3 made a positive change from considering observations facts to realizing that people can interpret observations differently. F3 credited this insight to working with other people throughout the course and seeing that they had different opinions about the same questions on in-class assessments. Two control group participants made externally influenced negative SUSSI changes. Both of these participants said that they made mistakes in their response on the post-SUSSI.

Table 19

Frequency of Positive and Negative External Influence Explanations

SUSSI NOS Aspect	Control Positive	Control Negative	Intervention Positive	Intervention Negative
(1) Observations and Inferences	1	2	—	—
(2) Change of Scientific Theories	—	—	—	—
(3) Laws vs. Theories	—	—	—	2
(4) Social and Cultural Influence	2	—	1	—
(5) Creativity and Imagination	1	—	—	—
(6) Scientific Methodology	3	—	—	—
Total	7	2	1	2

For the laws vs. theories NOS aspect, two intervention group participants used externally influenced arguments for negative SUSSI changes. Both shifted towards less informed views about the relationship between scientific theories and laws. Participant S1 said that their experiences in a psychology course they were taking contributed to this change. S8 attributed the change to things they had been observing in everyday life and thinking about how the world works.

One participant from each treatment group gave externally influenced explanations for changes they made on the social and cultural influence aspect. The control group participant made two positive changes. F14 shifted toward the view that culture determines both what and how science is conducted. F14 linked their shift for what science is conducted to learning about how various religions view science in another class. They credited the change in their view of culture influencing how science is conducted to learning about the male-dominated culture of science in the same class. Similarly, an intervention group participant, S7, positively shifted their views of the role of culture in science because of learning about the social contract in a sociology class they were taking.

There was only one externally influenced SUSSI change for imagination and creativity. A control group participant, F2, positively changed their views of the use of creativity for designing experiments. F2 gained this insight from observing the creativity of other students in the laboratory course associated with BIOS 1120.

Finally, two control group participants made positive changes in their views of scientific methodology. F7 realized that all science does not follow the same scientific method. When asked what caused their change in view, F7 noted learning about science done in other cultures like the Ottoman Empire in another class. Participant F2 shifted from thinking that scientific results are always correct when the scientific method is used to realizing that science can still be wrong regardless of method. F2 gained this insight again from their experiences in the laboratory course where they saw other students' experiments going wrong all the time.

Summary of SUSSI change explanation results. Historical and general course explanations were common across both the intervention and control groups. An important contrast was seen between the groups for historical explanations. Almost all of the historical explanations in the intervention group were associated with positive SUSSI Likert changes (7 of 8 explanations). The majority of historical explanations in the control group were linked to negative SUSSI Likert changes (7 of 11 explanations). One reason for this difference is likely the source that participants were drawing on for their historical explanations. In the intervention group, all of the historical explanations referenced the historical stories that were used for the instruction in that group. Control group participants mentioned examples from the course textbook for many of their historical explanations. The source that participants used is an important distinction because the historical stories introduced in the intervention group were designed to teach students about aspects of NOS. The minimal historical accounts in the textbook were likely not included with NOS in mind but rather to provide students with a general context. As a result, the textbook accounts may or may not have given accurate representations of NOS concepts.

Another important distinction between the two groups was the frequency of externally influenced explanations. These explanations were relatively common in the control group with six participants providing nine total externally influenced explanations. In the intervention group, externally influenced explanations were more rare with only three participants using them. It seems that control group participants were more likely to draw on their everyday life experience or outside coursework when responding to the SUSSI prompts.

Overall, the SUSSI change results showed that participants in the intervention group were more likely to use explanations based on the historical stories used in the course or their learning in the course more generally. Explanations given in the control group were more variable with participants drawing on the course textbook, general learning in the course, and experiences outside of the course. The SUSSI change results also provide support for the historical stories used in the intervention group being a positive influence on participant NOS views.

Connections between SUSSI quantitative results and participant explanations. The quantitative analysis done on the SUSSI Likert data, detailed in the results for research question 1, revealed several statistically significant differences from pre- to post-instruction across the two treatment groups. There was a significant improvement in mean composite SUSSI Likert scores for the imagination and creativity aspect for the intervention group. The control group had statistically significant decreases in composite scores for the observations and inferences and influence of society and culture SUSSI aspects. Interviewees from both groups had several explanations for changes in their SUSSI Likert responses that have relevance to the SUSSI quantitative findings.

Six intervention group interviewees discussed changes in their responses for the imagination and creativity SUSSI aspect. All six of the changes were positive changes where participants improved their NOS scores from pre- to post-instruction. When asked why they changed, three of these intervention group participants said that historical stories from the course influenced their thinking. One participant mentioned the Mendel story, one the moth story, and the other the historical stories in general as being important to their thinking.

The three participants that did not reference history all indicated that the instructor influenced their thinking. Two of these participants said that they remembered the instructor noting that creativity is an important aspect of conducting quality experiments. The other non-history participant said they answered that they were uncertain on the pre-SUSSI because they thought the instructor would want students to say imagination and creativity is not used in science. However, after learning about natural selection in the course they realized that imagination and creativity are indeed used in science.

In the control group, nine interviewees provided ten explanations for changes they made in their responses to SUSSI Likert items for the observations and inferences aspect. Six of the 10 explanations were for negative changes in SUSSI Likert scores. Three of the negative change explanations used historical examples related to Mendel and Darwin. These participants felt that the historical accounts of these scientists, likely from the textbook, showed them that observations are facts and that scientists always observe the same things because they are objective. The other three negative change explanations consisted of a general reference to the course and two participants that said they made mistakes when completing the post-SUSSI. The participant that gave the general course explanation said that readings from the textbook made them realize that observations are facts.

Five control group participants gave six explanations for changes they made on SUSSI items related to the influence of society and culture aspect. Three of the six explanations were for negative changes. Two participants referenced historical examples consistent with the course textbook in explaining their changes. Both of these changes were participants going from believing that society and culture influences science to thinking that it does not. The third participant stated that the instructor told the class that scientists put aside their culture when conducting their work.

All of these findings suggest that historical examples, the instructor, and the course textbook may have been important factors in the significant changes seen in the quantitative analysis of the SUSSI data. The positive changes in the intervention group interviewees' imagination and creativity scores were all linked to the historical stories used in the course or the course instructor. In the control group, the majority of negative changes for both the observations and inferences and the influence of society and culture SUSSI aspects were explained using historical examples. All of the historical examples mentioned by control group participants appeared to have been from the course textbook. Further evidence for the importance of historical examples for both groups will be discussed in the results for research question 4.

Explanations for changes in two-tier genetics instrument responses. Interview participants in both groups were also asked to explain any differences in their responses to two items on the pre- and post-instruction two-tier genetics instrument. Participants were provided with the items and their responses on the pre- and post-instrument. Only two items from the two-tier instrument were included in the interviews because of time considerations. The items provided to participants during the interviews were selected because their content was most relevant to the genetics instruction in the course. Additionally, the items were nearly identical between the pre- and post-version of the instrument. Unfortunately, participants in both groups had difficulty articulating their reasons for changing their two-tier genetics responses between testing occasions beyond general comments about the course. All of the explanations for changes on the two-tier genetics instrument were based on vague general comments about the course or the genetics unit being helpful. There was no difference between explanations for items where participants improved their score from pre- to post-instruction and those where scores got worse. These results provided support for the course genetics unit being helpful for genetics content. However, they did not provide information about why the genetics unit was helpful or any specific parts of the genetics unit that were particularly useful.

Research Question 4

Q4. What awareness do the participants have regarding the use of story and its associated narrative elements?

During the interviews, participants from both groups were asked about whether they interpreted the history used in the course as themes or as stories. Almost all of the intervention group participants responded that they interpreted the history as stories while the majority of control group participants responded that it was themes (Table 20).

Table 20

Frequency of Participant Interpretations of History

Control (*n*= 15)		Intervention (*n*= 12)	
Themes	Stories	Themes	Stories
10[a]	6[a]	1	11

[a]One participant said both stories and themes

Participants were subsequently asked why they interpreted the history as a story or a theme, whether they found historical stories or themes useful for learning about science, and whether the historical stories or themes made them feel more or less comfortable with learning about science (see Appendix C for interview script). Findings associated with these questions are described under two primary themes, (a) course historical content interpreted as a story/theme and (b) historical stories/themes are useful for learning about science.

Course historical content interpreted as a story/theme. Intervention group participants gave several reasons why they saw the history of science content included in the course as stories. The majority of these reasons fell under the general theme of the lives and work of scientists. Six out of the eleven intervention group participants felt that the instructor was telling a story because he would mention a specific scientist and talk about how they developed their theories. For example, participant S4 said, "Because I remember him talking about one or two specific scientists, and the way they used their research studies, and how they studied whatever it was that they studied at the time." S4 felt that because the instructor discussed a specific scientist and the way in which they studied a phenomenon, it made the history a story. Participant S8 made a similar point with reference to the Mendel story, "I think he would tell stories. Let me try and think. Like when he talked about Mendel and the pea plants, he'd tell stories and how he'd go about his experiments." Interestingly, these justifications regarding scientists' lives and work fit with the agency narrative element from Klassen's (2009) framework. The agency element says that stories consist of characters that make consequential decisions.

Other reasons given by intervention group participants included the presentation style of the instructor and that the historical material followed a story structure. Two participants thought that stories were being used in class because of how the history was presented in class. S13 commented:

> I saw it as a story because of the way he presented it in a story and then with the pictures and everything on the PowerPoint. Um him standing up there not reading from anything just knowing that like it was already in his head.

For S13, the visuals used during the stories and the preparation of the instructor made the history a story. Participant S6 recognized the basic story structure in the Moth story:

> Like he started from the beginning and talked about these moths and how they were all white, and then progressed and progressed about how scientists joined in. Scientists made their theories until the end of the theory that's the most accepted.

Participant S13's description of the moth story aligns with the minimal story structure proposed in the structure element of Klassen's (2009) framework. The structure element says that narratives have a beginning state, some event leading to change, and a final state.

The control group participants made analogous arguments for why they felt the history presented in their version of the course were not stories. Five of the fifteen control group interviewees mentioned that there was no story structure when asked why they did not think that the history included in class was stories. Several of these participants stated that a storyline or plot was missing, “I feel like he just said a general well for example this happens. If he like said well a full-on story plot, I feel like that would have made it more like a story (S11).” Other participants indicated that the history included was facts and not an account of past events, “It's like, this is how it's going to be and not ‘once upon a time’… he didn't put it in a context like that made it seem like a story (F1).” These justifications given by control group participants are consistent with the story structure and event-tokens elements from Klassen’s framework. The event-tokens element requires that stories include a series of events, taking place in a particular setting, that affect characters.

Some of the control group participants did state that the history included in the course was stories (six of fifteen interviewees). However, many of these participants gave examples or justifications that were not directly related to historical examples. Participant F6 talked about the structure of a typical day in class in which the instructor started off with background information, moved into the content for that day, and ended with the homework for the evening. F9 felt the way evolution content was presented as a progression of life through time made it a story. Other interviewees talked about the presentation style and enthusiasm of the instructor as making the history seem like stories.

These findings yield strong evidence that participants in both groups were able to recognize stories or the lack thereof. Many of the participants in the intervention and control groups gave justifications for their interpretations of the historical content in the class that were consistent with Klassen’s (2009) framework for historical stories. This indicates that the participants of this study at least had a general sense of when stories were being used. Interestingly, some control group participants were even interpreting the structure of the course and traditional science content as stories indicating that students are looking for stories in their learning whether consciously or not. Stories serving as a primary vehicle for learning is supported by the narrative literature (Klassen & Klassen, 2014; Noddings & Witherell, 1991).

Historical stories/themes are useful for learning about science. Questions asked during the interview about whether the historical stories/themes used in the course were helpful for learning about science gave valuable insights into how the stories/themes benefitted participant learning. Interviewees in both groups were asked a series of questions about whether the stories/themes were useful for course content, whether they gave insights about the practice of science, and if they made them more or less comfortable with learning about science. All of the intervention group participant said that stories were useful for learning about science. The majority mentioned either the Mendel story (four of twelve) or the moth story (four of twelve). Of the participants that mentioned the Mendel story, one said that it provided insights about how scientists do observations (S4). Two participants, S1 and S11, noted that the Mendel story showed them scientists change their minds about how things work and the way experiments should be conducted. S11 said:

> Anyway, the founding father behind genetics kind of like helped us explain how we look where we are today. It kind of like gives you the idea of, hey, what if I try this on a smaller scale to get a better understanding of a bigger picture of it?

Participants S1 and S11 both seemed to think the Mendel story was an example of a scientist breaking with previous work and trying something new.

Three of the intervention group participants that referenced the moth story said that it was helpful for learning and remembering course content related to natural selection. S5 compared the story approach to natural selection to a more traditional approach, “And it probably just would have been like, OK, here's something else I need to memorize, instead of, here's a fun anecdote to try to make it stick in there in a different way outside of the box.” For these participants, the moth story provided an interesting example of natural selection instead of just the facts of the concept. Three of the interviewees that mentioned the moth story said that it gave them insights about how science works. Both S7 and S13 said that the moth story showed them that many scientists are involved in studying a given phenomenon. S7 also added that it gave them a sense of the amount of trial and error in science. S6 stated that the moth story demonstrated that experiments are not the only way to generate scientific knowledge. In general, these participants indicated that the moth story was helpful for making the natural selection content memorable and for providing a better understanding of the flexibility and messiness of the scientific process.

Nine of the intervention group participants also discussed why they thought stories more generally were helpful to their learning of science. Five of the nine maintained that stories provide real world examples of science which in turn makes content easier to remember. S5 reasoned that the stories made learning about science feel more personal and relatable:

> It made it feel more personal. I don't know. Like little stories, I don't know. That seemed to click with me, the stories, making it seem more like not just these are the scientific-- that this is a story that you are reading that just happens to be true. It's something that happened that you need to learn. I liked that.

Two other participants, S8 and S13, talked about how the stories compared favorably to learning from a textbook. S8 stated, "Because instead of just maybe looking into a textbook and just reading some bland stuff, it kind of helped you get a real world-- I'd say, like, where do we use this every day? What people experience this every day?" For these participants, the stories were more interesting and easier to understand than the textbook.

Another interesting theme that emerged from the intervention group interviews was that several participants (four of twelve) considered stories easier to relate to and more interesting for students that do not particularly like science. S6 noted that for many people science classes are general education requirements that they need to get done but do not particularly enjoy. Stories can be helpful for these types of students because, "I feel like kind of adding something that's not just lecture, lecture, lecture adds more to the class. It made me pay attention more." Similarly, S4 said that often science courses can be a boring stream of facts that did not trigger their mind. Anything that can be added to a course that, "grabs students' attention in a different way", is a positive.

Control group responses to the utility of historical themes/stories questions largely aligned with the findings from the intervention group. The majority of control group participants (nine of fifteen) said that historical themes/stories were helpful for learning about science. It is important to point out that the historical examples mentioned by control group participants were all likely from the textbook as the instructor used minimal history during the control group instruction. Control group interviewees commonly stated that the historical themes/stories provided useful background information that made content easier to learn (four of nine), gave them insights about how science works (three of nine), and served as real world examples that made content easier to relate to and learn (five of nine). Interviewees made general comments about how the historical themes/stories gave them important background information that prepared them for upcoming lessons. F2 went into more detail and said that historical themes can make science less intimidating to learn about because, "if you put a face to it then it's like, oh, some guy figured this out and it's pretty cool." For F2, it was comforting to learn that people actually worked on and developed scientific ideas.

Two of the three control group participants that said they gained insights about how science works talked about learning that science changes over time. For F7 and F11, learning about the history of evolutionary thought before and after Darwin illustrated how scientific theories change over time. The third participant said that learning about the history of the study of DNA made them realize that science is a race towards the next discovery.

Similar to the intervention group, control group participants also said that the historical stories made it easier for non-science students to learn science content. F6 and F12 both felt that the history presented was stories and that it made the content of the course more understandable. F6 said, "…I have never been super into science so making it more of like a history lesson just like made I don't know more of a story less about facts." These participants were less intimidated by science when it was presented with historical context.

Finally, it should be noted that four control group participants did not think that the historical themes/stories were useful for learning about science. A range of reasons were given including preference for rote learning, being uninterested in history or finding it difficult to remember, and feeling the history was not well connected to course content.

Summary. Overall, the results for research question 4 revealed several insights about the dispositions of the students involved in the current study toward historical stories. These results also gave some understanding of how the historical content included in both treatment groups was experienced by participants. The majority of study participants across both groups were able to recognize and or articulate an understanding of stories. This is supported by many participant interpretations of stories aligning with narrative elements from Klassen's (2009) framework such as agency and event-tokens. Additionally, intervention group participants were able to discuss numerous reasons why the historical stories were beneficial to their learning. A large majority of the intervention group stated that the stories gave them insights into how science works suggesting that stories were an important influence on their NOS understanding. It is important to note that a majority of the intervention group also mentioned either the Mendel story or the moth story specifically. The specific mention of these stories suggests the Mendel story and the moth story influenced NOS views in the intervention group. Many participants across both groups also talked about how historical accounts made course content easier to understand and remember. Additionally, there were several participants across both groups that said the historical accounts made the content more accessible specifically for students who do not like science. These findings suggest support for the use of history and in particular historical stories in the science classroom as a tool for teaching students about both traditional science content and NOS. The next section discusses the conclusions, limitations, and implications of the research findings as a whole.

CHAPTER V

CONCLUSION

The primary focus of this study was to investigate the influence of the introduction of historical stories on undergraduate students' understanding of NOS concepts and traditional genetics content. The current chapter will tie together all of the previous chapters to bring forth the key findings of the study. Chapter one overviewed the difficulty learners have with learning concepts related to NOS. NOS views have shown to be resistant to change. Incorporating the history of science has been shown to be a promising approach for overcoming the challenges learners have with NOS. However, the history of science has been inconsistently used in the science education literature. Defining what using the history of science in the science classroom means is important for determining its utility for teaching about NOS. One way history of science can be defined is through stories. Klassen (2009) laid out a framework for creating historical stories for use in the science classroom. The framework has the potential to provide a more standardized approach to the use of history. This is not to say that there should be one way to use history but rather that there is a practical need for consistent definitions in order to move the research forward in this area.

Chapter two took these ideas forward and began by reviewing the literature related to the teaching and learning of NOS. Researchers have established that an explicit and reflective approach should be used that is contextualized within real world examples. The chapter then moves on to discuss research that has been done on the role of the history of science in science education. Several theoretical arguments for the history of science were presented along with empirical studies done on the effectiveness of using the history of science to improve NOS views and science content. The results of these empirical studies have been inconsistent. As stated previously, one reason for this is that the implementation of history in empirical studies has not been consistent. The chapter then considers historical stories as an approach to using history of science in the science classroom. This part of chapter two primarily focuses on the historical story framework proposed by Klassen (2009). The framework provides a clear set of elements that should be included in a story that draws upon narrative research. However, to date there has been little empirical research done on the effectiveness of Klassen's framework.

The current study is intended to begin to address this gap and chapter 3 details the approach taken to determine the effectiveness of historical stories for improving NOS views and genetics content knowledge. The story selected for the current study, based on the life and work of Gregor Mendel, was chosen because it includes all of the elements from Klassen's (2009) framework. The rest of chapter 3 describes how a control and intervention group were used to determine the effectiveness of the Mendel story. Methods used for collecting and analyzing data from the SUSSI, a two-tier genetics instrument, and interviews are described. These analyses were used to determine what effects the Mendel story had on the NOS views and genetics content understanding of intervention group participants. Chapter four discusses the findings of the current study. Chapter five considers the gaps in the research identified in chapter two and considers how the current study addresses these gaps and the significance of the findings. Limitations and implications of the findings are also considered.

Research Gaps

As was noted in chapters one and two, an understanding of NOS is a longstanding and uncommonly attained goal of science education. NOS is considered an important part of scientific literacy. Despite significant effort to improve NOS understanding of both students and science instructors, both groups continue to struggle (Lederman & Lederman, 2014). One reason that struggles continue with NOS understanding is that science instructors often do not want to take time to teach NOS. Many instructors see NOS instruction as taking away from time for traditional science content (Clough, 2006). One way that researchers have sought to address issues with NOS teaching and learning is through creating and conducting research on explicit and reflective instructional approaches using contextualized examples of NOS concepts.

One approach to contextualized explicit and reflective NOS instruction that is advocated by many researchers is using examples from the history of science to teach learners about NOS. History of science is seen as promising because among many other reasons it allows for a natural mixing of NOS and traditional science content (Clough, 2006). Several empirical studies have been done to determine the effectiveness of historical approaches to teaching NOS with variable results. Abd-El-Khalick and Lederman (2000) found that history of science courses had a minimal effect on the NOS views of undergraduates and pre-service teachers. In contrast, studies by Kim and Irving (2010) and Rudge et al. (2014) among others have shown positive effects on NOS views from historical interventions. A likely contributing factor to the variability in findings related to the effectiveness of the history of science for improving NOS is that the way history is delivered across studies is also variable.

Recently some science education researchers have begun advocating the use of historical stories in science instruction (Metz et al., 2007; Klassen, 2009, Klassen & Klassen, 2014). Klassen (2009) proposed a framework for creating historical stories for the classroom that calls for ten elements that should be included in a historical story. The framework could be used as one way to standardize the use of history in research for the purposes of improving comparability between studies. However, there is currently a lack of quality empirical research supporting the use of historical stories based on the Klassen (2009) framework for improving students NOS views or science content understanding outside of Fulford (2016).

Based upon the brief review of literature discussed in chapters one and two above, there are three primary gaps in the literature that the current study sought to address. First, the variability in findings regarding the use of the history of science for improving undergraduate students understanding of NOS. Second, the lack of empirical evidence regarding the use of historical instructional approaches to address both NOS and science content understanding together. Third, the lack of empirical evidence regarding the use of historical stories based on Klassen's (2009) framework for improving NOS and science content understanding. The following section discusses these gaps in relation to the findings of the current study.

Use of the history of science to improve NOS views. As was stated above, previous studies have reported conflicting results about whether using the history of science affects student nature of science views. The current study provides additional support for the idea that historically based instructional approaches can improve undergraduate student understanding of NOS concepts. The MANOVA analysis revealed a statistically significant increase in composite scores for the imagination and creativity aspect of the SUSSI for the intervention group from pre- to post-instruction. No significant increases in SUSSI composite scores were seen for the control group. Additionally, both treatment groups had relatively large sample sizes, comparable demographics, and consistent instruction with the exception of the historical stories in the intervention group. All of these results provide quantitative evidence that the historical stories positively influenced the NOS views of the undergraduate student participants related to imagination and creativity in science.

In addition to the purely quantitative data, there was support for the historical stories positively influencing participant NOS views in the interview data. First, out of twenty-four total SUSSI change explanations discussed during interviews with intervention group participants, eight were explained with reference to the historical stories. Seven of these eight historical explanations were associated with positive changes in SUSSI scores. Similar results were also seen specifically related to participant explanations for changes in their responses to the imagination and creativity SUSSI aspect. All six of the changes for the intervention group were positive. Of these six, three were explained in relation to the historical stories including one participant that mentioned the Mendel story, one that mentioned the moth story, and one that talked about the stories in general. When contrasted with the control group, these results indicate that the historical stories had a positive influence on intervention participants NOS views. Finally, when asked about their perceptions of the history used in class and whether it gave them insights about how science works the response from intervention group participants was overwhelmingly positive. Eleven of the twelve intervention group participants interpreted the history used as stories. The majority also said that the stories gave them insights about how science works with four mentioning the Mendel story and four mentioning the moth story. These qualitative findings provide evidence that the historical stories played an important role in influencing the NOS views of the intervention group. Together the quantitative and qualitative data strongly support that the historical stories positively influenced participants' NOS views.

Use of the history of science to improve both NOS and science content. As previously discussed in chapter two, there is little empirical research showing support for historical approaches leading to improvements in both NOS and traditional science content. This is particularly true in the context of genetics content. One of the few available studies by Kim and Irving (2010) reported promising results but had issues with the validity and reliability of the instruments used. The current study provides some support for historical stories positively influencing the genetics content understanding of the intervention group participants.

Quantitative analysis using McNemar's test revealed statistically significant increases in the number of correct responses for four items on the genetics two-tier instrument for the intervention group. Importantly all four items were directly related to Mendelian genetics which was one of the learning goals for the genetics unit in the course. In contrast, significant increases were seen in two items for the control group. Since genetics instruction was the same between the two groups with the exception of the Mendel story, these results suggest that the Mendel story helped intervention group participants perform better in regard to Mendelian genetics content knowledge.

The interview data did not speak directly to the results seen on the genetics two-tier instrument. However, the majority of intervention group participants said that the historical stories made course content easier to remember. Many control group participants made similar comments regarding history. Four participants in the intervention group talked about the historical stories making course content more interesting and relatable for students that do not like science. These results suggest that the Mendel story could have helped to engage students more in advance of the subsequent genetics unit. Considering the evidence presented in the previous section regarding NOS views, the current study provides support for historical instructional approaches being able to address both NOS and traditional science content.

Empirical support for Klassen's story framework. There are currently only two empirical studies that provide support for Klassen's story framework. One of studies done by Klassen himself (2009) explored a story about Louis Slotin created using the framework. The data collection and analysis was quite limited because student questions and ideas about the purpose of the study were the only things collected. Data was analyzed by organizing student responses in *apriori* categories based on the stem word of the student questions and the topic of their suggested story purposes. These data were used as a measure of the stories ability to engage students. In the absence of additional data sources, the Klassen (2009) study does not provide much information on the utility of the story framework. The other study by Fulford (2016) currently provides the highest quality empirical data on the Klassen (2009) story framework. Fulford (2016) looked at the effects of a historical story about Industrial Melanism developed using the framework. Data collected through surveys and interviews indicated the story was effective in decreasing student misconceptions about natural selection. The current study moves beyond the Klassen (2009) study and builds off of the Fulford (2016) study. The evidence in the previous two sections support the potential for stories developed using the Klassen (2009) framework to positively influence NOS views and traditional science content. This study adds to the literature by being one of the first to test the effects of a story, with Klassen's (2009) elements, on NOS views of undergraduate students.

There is also additional support for Klassen's story framework in the interview data. Many of the intervention group participants' ideas for why they thought the history presented was a story were consistent with Klassen's elements such as agency, event-tokens, and the minimal story structure. Similarly, many control group participants thought the history included was not stories because of a lack of minimal story structure. These results speak to the value of the framework for resonating with learners because at least some of the elements were things participants recognized independently. Additional support for this point came from a control group participant mentioned in chapter 4 that used the beginning, middle, and end structure of a typical day in class as an example of stories. The results are also in alignment with Fulford's (2016) study where all of Klassen's elements were identified by the participants as a whole.

Limitations

There were limitations that emerged from this study based on the design, instrumentation used, and the execution. Both the quantitative and qualitative portions of the study had limitations.

Quantitative. SUSSI open response data was collected from each of the participants that completed the SUSSI instrument. However, this data ended up not being useful after analysis for two primary reasons. One, participants tended not to write very much and had difficulty articulating examples to support their points. This in turn led to the second limitation of the open response data which was the scoring scheme suggested by the authors of the SUSSI. For most of the open response items, the majority of participant responses ended up being coded a 2 or "transitional". Many times, the transitional score seemed to be assigned for arbitrary reasons. For example, the society and culture open response item asks participants to "explain how society and culture affect or do not affect science." The scoring scheme only grants a 3 or "informed view" if participants say in their answer that society affects what and how science is conducted. If the participant only says what and not how then they automatically receive a 2. Since the prompt gives participants no way of knowing that it is looking for both it is not surprising when both are not included resulting in many responses defaulting to a score of 2. Since the scale has difficulty differentiating between informed and less informed it makes it difficult to use for meaningful analysis.

Qualitative. Two primary limitations came out of the qualitative portion of the study. First, it was difficult to get participants to talk about their responses on the genetics two-tier instrument. It was intended that data similar to that collected for the SUSSI Likert responses would be collected for the genetics two-tier instrument. Many participants either seemed to not know how to answer questions about their genetics item responses or seemed uncomfortable with talking about their answer with the interviewer. The discomfort for some participants seemed to come from feeling like they had given the wrong answer on the instrument. As a result, qualitative data on why participants may have changed their responses on the genetics two-tier instrument was limited.

Additionally, the interviewers approach may have caused some data to be missed. Upon reviewing the transcripts some opportunities for follow up questions were not taken. Sometimes these were not taken because of the interviewer reading cues from the interviewee. However, there are certainly some that were simply missed due to a lack of experience.

Overall study design. The presence of a historical story in addition to the Mendel story in intervention group makes interpreting the results of the study more difficult. In particular, the differences seen in the SUSSI results, the two-tier genetics results, and in the interviews cannot be as clearly attributed to the Mendel story except for the cases when participants talked about the Mendel story directly. The moth story could not have been avoided because data collection for this study was done as part of a larger study involving historical stories. While the additional story reduces the claims that can be made directly regarding the Mendel story they do not diminish claims made about Klassen (2009) framework stories more generally.

Implications

This section includes reflections on the research findings and the limitations of the study. It provides suggestions on how the study could be improved and directions for future research.

SUSSI. The SUSSI was selected because it is a recently developed, valid and reliable instrument that was appropriate for the planned sample size for this study. However, the open-ended questions included on SUSSI did not end up providing useful data for this study. Researchers conducting future studies with the SUSSI should consider revising the open-ended question prompts and or the scoring scheme. Adding finer grain distinctions between scores on the SUSSI open-ended item scoring scheme would help differentiate between participants with more or less informed NOS views. This being said, whether the scoring scheme is edited or not the questions need to be brought into better alignment with the expectations of the scoring scheme. Making these changes would lead to more useful data from these items.

The issues with the open-response items and the abstract nature of the Likert item prompts suggest that a new instrument for studying NOS views should be developed. The new instrument should be contextualized within historical and modern examples of science. Providing study participants with real world examples would likely help to mitigate issues with students not understanding some of language on the SUSSI. For example, some students seemed to struggle with the use of the word imagination on the imagination and creation SUSSI items. Historical and modern examples of science would be easier for participants to interpret than non-contextualized prompts like the Likert items on the SUSSI. Starting points for a new instrument include literature based on the Argumentative Resource framework (Allchin, 2011; Sandoval et al., 2005; Deng et al., 2011).

Two-tier genetics instrument. Similar to the SUSSI, the two-tier genetics instrument was selected because it is a recently developed, valid and reliable instrument that was aligned with the genetics content taught in the course. While the instrument was not designed with qualitative interview data in mind, the poor interview data collected related to the genetics two-tier instrument was still disappointing. More directed interview prompts would likely have helped participants provide better responses. Also, taking care to ensure participants that they were not being interviewed because their answers are wrong may have put them more at ease.

Mendel story. The Mendel story was intended to introduce the genetics unit and grab students' attention prior to their learning course content. One way the story could be improved is through adding some type of hands-on activity for the students to do in the midst of the story. Adding an activity to the story for this study was not feasible because of the class size and the amount of class time (50 minutes). However, computer simulations tie in nicely with discussions of Mendel (Williams & Rudge, 2015;2016).

Future research. In order to continue advancing this line of research, future researchers should consider how stories have been used in science education studies to this point. Particular attention should be paid to the main protagonists that have been used in most studies. The majority of historical science stories that have been used in the literature have focused on white male scientists. Important insights could be gained comparing learner responses to stories centered on women and minorities in comparison to the more typical white male focused stories. The current study showed that a short intervention involving a historical story with Klassen's elements used as a door opener can influence NOS views and science content understanding. It would be helpful to see how Klassen framework stories that were more directly integrated with science content affect learners NOS and science content understandings. Learning more about how the level of story and science content integration affects NOS understanding would be particularly interesting given the of lack of studies on NOS and Klassen framework stories. An in-depth qualitative study would be particularly useful for developing a deeper understanding of how learners are interacting with historical stories and how they influence their understandings. Finally, future research should also focus on a contextualized instrument for evaluating NOS views.

REFERENCES

Abd-El-Khalick, F., & Lederman, N. 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. doi: 10.1002/1098-2736(200012)37:10<1057::AID-TEA3>3.0.CO;2-C

Aikenhead, G. S., & Ryan, A. G. (1992). The Development of a New Instrument: "Views on Science—Technology—Society" (VOSTS). *Science Education*, *76*(5), 477–491. doi:10.1002/sce.3730760503

Akerson, V. L., Abd-El-Khalick, F., & Lederman, N. G. (2000). Influence of a reflective explicit activity-based approach on elementary teachers' conceptions of nature of science. *Journal of Research in Science Teaching*, *37*(4), 295–317. doi:10.1002/(SICI)1098-2736(200004)37:4<295::AID-TEA2>3.0.CO;2-2

Allchin, D. (2003). Scientific Myth-Conceptions. *Science Education*, *87*(3), 329–351. doi: 10.1002/sce.10055

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education*, *95*(3), 518–542. doi: 10.1002/sce.20432

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy: A Project 2061 report*. New York: Oxford University Press.

American Association for the Advancement of Science. (2009). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Appleton, K. (1997). Analysis and description of students' learning during science classes using a constructivist-based model. *Journal of Research in Science Teaching*, *34*(3), 303–318. doi: 10.1002/(SICI)1098-2736(199703)34:3<303::AID-TEA6>3.0.CO;2-W

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, *51*(5), 267. doi: 10.1037/h0046669

Bell, R., & Lederman, N. (2003). Understandings of the nature of science and decision making on science and technology based issues. *Science Education*, *87*(3), 352–377. doi: 10.1002/sce.10063

Biological Sciences Curriculum Study [BSCS]. (1962). *Processes of science test*. New York: The Psychological Corporation.

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York, NY: Sage.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Byrne, B. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294–320. doi: http://dx.doi.org/10.1177%2F0049124113500475

Central Association of Science and Mathematics Teachers (1909). A consideration of the principles that should determine the courses in biology in the secondary schools. *School Science and Mathematics, 9,* 241–247.

Chen, S. (2006). Development of an instrument to assess views on nature of science and attitudes toward teaching science. *Science Education*, *90*(5), 803–819. doi: 10.1002/sce.20147

Clough, M. P. (2006). Learners' Responses to the Demands of Conceptual Change: Considerations for Effective Nature of Science Instruction. *Science & Education*, *15*(5), 463–494. doi: 10.1007/s11191-005-4846-7

Clough, M. P. (2011). The story behind the science: Bringing science and scientists to life in post-secondary science education. *Science & Education*, *20*(7), 701–717. doi: 10.1007/s11191-010-9310-7

Clough, M. P., Herman, B. C., & Smith, J. A. R. (2010). Seamlessly teaching science content and the nature of science: Impact of historical short stories on post-secondary biology students. Paper presented at the Association for Science Teacher Education (ASTE) National Conference, Sacramento, CA.

Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.

Creswell, J. W., Clark, V. L. P., Gutmann, M., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 209–240). Thousand Oaks, CA: Sage.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Putnam.

Deng, F., Chen, D., Tsai, C., & Chai, C. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, *95*(6), 961–999. doi:10.1002/sce.20460

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

Duit, R., & Treagust, D. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671–688. doi:10.1080/09500690305016

Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education*, *85*(5), 554–567. doi:10.1002/sce.1023

Fulford, J.M. (2016). *Assessing the Impact of Historical Story Telling on Student Learning of Natural Selection* (Doctoral dissertation). Retrieved from http://scholarworks.wmich.edu/dissertations/1422

Garrison, J. (1994). Realism, Deweyan pragmatism, and educational research. *Educational Researcher*, *23*, 5–14. doi:10.3102%2F0013189X023001005

Gericke, N., & Hagberg, M. (2007). Definition of historical models of gene function and their relation to students' understanding of genetics. *Science & Education*, *16*(7), 849–881. doi: 10.1007/s11191-006-9064-4

Gericke, N., & Smith, M. (2014). Twenty-First-Century Genetics and Genomics: Contributions of HPS-Informed Research and Pedagogy. In M. R. Matthews (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 423-467). doi: 10.1007/978-94-007-7654-8_15

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*, 59–82. doi: http://dx.doi.org/10.1177%2F1525822X05279903

Hadzigeorgiou, Y., Klassen, S., & Klassen, C. (2012). Encouraging a "romantic understanding" of science: The effect of the Nikola Tesla story. *Science & Education*, *21*(8), 1111–1138. doi: 10.1007/s11191-011-9417-5

Hanuscin, D. L., Akerson, V. L., & Philipson-Mower, T. (2006). Integrating nature of science instruction into a physical science content course for preservice elementary teachers: NOS views of teaching assistants. *Science Education*, *90*(5), 912–935. doi: 10.1002/sce.20149

Hatcher, L., & Stepanski, E. J. (1994). *A step-by-step approach to using the SAS system for univariate and multivariate statistics*. Cary, NC: SAS Institute.

Hodson, D. (2014). Nature of Science in the Science Curriculum: Origin, Development, Implications and Shifting Emphases. In M. R. Matthews (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 911–970). doi: 10.1007/978-94-007-7654-8_28

Howard, P. J. (2000). *The owner's manual for the brain: Everyday applications from mind-brain research* (2nd ed.). Austin, TX: Bard Press.

Howe, E. M., & Rudge, D. W. (2005). Recapitulating the History of Sickle-Cell Anemia Research. *Science & Education*, *14*(3), 423–441. doi:10.1007/s11191-004-1996-y

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage Publications, Inc.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. doi:10.1080/10705519909540118

Hungerford, H. & Walding, H. (1974). The modification of elementary methods students' concepts concerning science and scientists. Paper presented at the Annual Meeting of the National Science Teachers Association.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. doi:10.3102%2F0013189X033007014

Khishfe, R. (2012). Relationship between nature of science understandings and argumentation skills: A role for counterargument and contextual factors. *Journal of Research in Science Teaching*, *49*(4), 489–514. doi: 10.1002/tea.21012

Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching*, *39*(7), 551–578. doi:10.1002/tea.10036

Klassen, S. (2009). The construction and analysis of a science story: A proposed methodology. *Science & Education*, *18*(3-4), 401–423. doi:10.1007/s11191-008-9141-y

Klassen, S. (2010). The relation of story structure to a model of conceptual change in science learning. *Science & Education*, *19*(3), 305–317. doi:10.1007/s11191-009-9212-8

Klassen, S., & Klassen, C. (2014). Science teaching with historically based stories: Theoretical and practical perspectives. In M. R. Matthews (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 1503–1529). doi: 10.1007/978-94-007-7654-8_47

Kim, S., & Irving, K. (2010). History of Science as an Instructional Context: Student Learning in Genetics and Nature of Science. *Science & Education*, *19*(2), 187–215. doi: 10.1007/s11191-009-9191-9

Leach, J., Hind, A., & Ryder, J. (2003). Designing and evaluating short teaching interventions about the epistemology of science in high school classrooms. *Science Education, 71*, 117–134. doi: 10.1002/sce.10072

Lederman, N. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 831–879). Mahwah, NJ: Erlbaum.

Lederman, N., Abd-El-Khalick, F., Bell, R. & Schwartz, R. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521. doi: 10.1002/tea.10034

Lederman, N., Bartos, S., & Lederman, J. (2014). The development, use, and interpretation of nature of science assessments. In M. R. Matthews (Ed.), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 971–997). doi:10.1007/978-94-007-7654-8_29

Leech, N., & Onwuegbuzie, A. (2009). A typology of mixed methods research designs. *Quality & Quantity*, *43*(2), 265–275. doi:10.1007/s11135-007-9105-3

Liang, L. L., Chen, S., Chen, X., Kaya, O. N., Adams, A. D., Macklin, M., & Ebenezer, J. (2008). Assessing Preservice Elementary Teachers' Views on the Nature of Scientific Knowledge: A Dual-Response Instrument. *Asia-Pacific Forum on Science Learning and Teaching*, *9*, 1–20.

Liang, L., Chen, S., Chen, X., Kaya, O., Adams, A., Macklin, M., & Ebenezer, J. (2009). Preservice Teachers' Views About Nature of Scientific Knowledge Development: An International Collaborative Study. *International Journal of Science and Mathematics Education*, *7*(5), 987–1012. doi:10.1007/s10763-008-9140-0

Lin, H., & Chen, C.C. (2002). Promoting preservice chemistry teachers' understanding about the nature of science through history. *Journal of Research in Science Teaching*, *39*(9), 773–792. doi:10.1002/tea.10045

MacCallum, R., & Browne, M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. doi:10.1037/1082-989X.1.2.130

Marshall, C., & Rossman, G. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Matthews, M. R. (1994). *Science Teaching: The Role of History and Philosophy of Science*. New York, NY: Routledge.

Matthews, M. (1997). Introductory comments on philosophy and constructivism in science education. *Science & Education*, *6*, 5–14. doi:10.1023/A:1008650823980

McComas, W. F. (2004). Keys to Teaching the Nature of Science. *Science Teacher*, *71*(9), 24–27. Retrieved from https://search.proquest.com/docview/214618434?pq-origsite=gscholar

McComas, W. (2010). The History of Science and the Future of Science Education. In P. V. Kokkotas, K. S. Malamitsa, & A. A. Rizaki (Eds.), *Adapting Historical Knowledge Production to the Classroom* (pp. 37–53). doi: 10.1007/978-94-6091-349-5_3

McComas, W., & Olson, J. (1998). The nature of science in international science education standards documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41-52). Dordrecht, The Netherlands: Kluwer Academic Publishers.

McElhinny, T., & Dougherty, M. (2014). The status of genetics curriculum in higher education in the United States: goals and assessment. *Science & Education*, *23*(2), 445–464. doi: 10.1007/s11191-012-9566-1

Metz, D., Klassen, S., McMillan, B., Clough, M., & Olson, J. (2007). Building a foundation for the use of historical narratives. *Science & Education*, *16*(3), 313–334. doi: 10.1007/s11191-006-9024-z

Miller, M. C. D., Montplaisir, L. M., Offerdahl, E. G., Cheng, F. C., & Ketterling, G. L. (2010). Comparison of Views of the Nature of Science between Natural Science and Nonscience Majors. *CBE-Life Sciences Education*, *9*(1), 45–54. doi:10.1187/cbe.09-05-0029

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*(4), 405–424. doi: 10.1002/(SICI)1098-237X(199707)81:4<405::AID-SCE3>3.0.CO;2-G

Moore, R., & Sutman, F. (1970). The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching*, 7, 85–94. doi: 10.1002/tea.3660070203

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

Noddings, N., & Witherell, C. (1991). Epilogue: Themes remembered and foreseen. In C. Witherell & N. Noddings (Eds.), *Stories lives tell* (pp. 279–280). New York, NY: Teachers College Press.

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity*, *41*(2), 233–249. doi:10.1007/s11135-006-9000-3

Onwuegbuzie, A. J., & McLean, J. E. (Ed). (2003). Expanding the Framework of Internal and External Validity in Quantitative Research. *Research in the Schools*, *10*, 71–89.

Onwuegbuzie, A., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 351–383). Thousand Oaks, CA: Sage.

Park, H., Nielsen, W., & Woodruff, E. (2014). Students' conceptions of the nature of science: perspectives from Canadian and Korean middle school students. *Science & Education*, *23*(5), 1169–1196. doi:10.1007/s11191-013-9613-6

Posner, G., Srike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*, 211-227. doi: 10.1002/sce.3730660207

Researchware, Inc. (2015). HyperRESEARCH [Computer Software]. Retrieved from http://www.researchware.com

Richardson, J. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. doi:10.1016/j.edurev.2010.12.001

Rorty, R. (2000). Pragmatism. *International Journal of Psychoanalysis*, *81*(4), 819–823.

Rudge, D. W., Cassidy, D. P., Fulford, J. M., & Howe, E. M. (2014). Changes Observed in Views of Nature of Science During a Historically Based Unit. *Science & Education*, *23*(9), 1879–1909. doi:10.1007/s11191-012-9572-3

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*(5), 561–580. doi:10.1007/s11191-007-9088-4

Sadler, T., & Fowler, S. (2006). A threshold model of content knowledge transfer for socioscientific argumentation. *Science Education*, *90*(6), 986–1004. doi:10.1002/sce.20165

Sadler, T., & Zeidler, D. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, *89*, 71–93. doi:10.1002/sce.20023

Sandoval, W. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, *12*, 5–51. doi:10.1207/S15327809JLS1201_2

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

Schwirian, P. M. (1968). On measuring attitudes toward science. *Science Education*, *52*, 172–179. doi: 10.1002/sce.3730520212

Steiger, J. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7*(2), 149–162. doi:10.1207/S15328007SEM0702_1

Stice, G. (1958). *Facts about science test.* Princeton, NJ: Educational Testing Service.

Tao, P. (2003). Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories. *International Journal of Science Education*, *25*(2), 147–171. doi:10.1080/09500690210126748

Treagust, D. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169. doi:10.1080/0950069880100204

Tsui, C., & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, *32*(8), 1073–1098. doi:10.1080/09500690902951429

Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., & Beauchamp, C. (2001). Analyzing interview data: The development and evolution of a coding system. *Qualitative Sociology*, *24*(3), 381–400. doi:10.1023/A:1010690908200

Williams, B., Clough, M. P., Stanley, M., & Colbert, J. T. (2010). Creativity and discovery: The work of Gregor Mendel. Retrieved from http://www.storybehindthescience.org/biology.html

Williams, C., & Rudge, D. (2015). Mendel and the Nature of Science. *The American Biology Teacher*, *77*(7), 492–499. doi:10.1525/abt.2015.77.7.3

Williams, C., & Rudge, D. (2016). Emphasizing the History of Genetics in an Explicit and Reflective Approach to Teaching the Nature of Science. *Science & Education*, *25*(3-4), 407–427. doi:10.1007/s11191-016-9821-y

Wilson, L. (1954). A study of opinions related to the nature of science and its purpose in society. *Science Education*, 38(2), 159–164. doi:10.1002/sce.3730380209

Zeidler, D., Walker, K., & Ackett, W. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education*, *86*(3), 343–367. doi:10.1002/sce.10025

APPENDIX A

HSIRB Approval Forms

WESTERN MICHIGAN UNIVERSITY

Human Subjects Institutional Review Board

Date: August 7, 2014

To: David Rudge, Principal Investigator
Cody Williams, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 14-08-01

This letter will serve as confirmation that your research project titled "Learning in a Biology Content Lecture Course" 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: August 6, 2015

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: Cody Williams

Title of Study: **Learning in a Biology Content Lecture Course**

You have been invited to participate in a research project titled "Learning in a Biology Content Lecture Course" 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 different ways of introducing biological content on student learning and attitudes towards science.

Who can participate in this study?
Any student enrolled in Dr. Rudge's section of BIOS 1120 can participate in this study.

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

What is the time commitment for participating in this study?
Participation consists in giving Dr. Rudge permission to use any artifact you create as a result of required classroom activities (pre-assessment surveys, reflection essays, exams) for research purposes after final grades are submitted. *Please note that in the event you choose to do so, all identifying information will be removed.* For most subjects, there is no time commitment over and beyond the time you are spending reviewing and signing this consent form. Some subjects (who specifically indicate below that they wish to be contacted) may be asked to participate in a 30 minute interview outside of class.

What will you be asked to do if you choose to participate in this study?
By choosing to participate in this study, you are giving Dr. Rudge permission to use any artifact you create as a result of required classroom activities (completed pre-assessment surveys, reflection essays, exams) for research purposes. If you specifically indicate below that you wish to be contacted, you may also be asked to participate in an interview regarding your impressions of the course.

What information is being measured during the study?
Your responses to the assessment instruments and other classroom artifacts will be collected and coded. The interviews will be recorded using a digital recorder and transcribed. The responses to the instruments and the interviews will then be analyzed to determine the impact of the course format on student learning and also attitudes towards science.

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. There are no known risks. However, in order to minimize 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. Your instructor will not be told until after final grades have been submitted who agreed to participate in this study. 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. Data collection and analysis are being done to improve the course for future students. This study will also potentially benefit the field of science education by adding to our understanding of how students learn science and how attitudes about science may be affected by instructional practice.

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 for the purposes of the interview.

Is there any compensation for participating in this study?
There will be extra credit available to subjects participating in the interview portion of this study.

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 assessments and redacted quotes from the interviews will be included in all forms of dissemination.

Risks and Cost to and Protections for Subjects

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.

Western Michigan University
H.S.I.R.B.
Approved for use for one year from this date: 3

AUG 07 2015

x
HSIRB Office

Should you have any questions prior to or during the study, you can contact the student investigator, Cody Williams at 269-387-5398 or cody.t.williams@wmich.edu. 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:____________________________

APPENDIX B

Recruitment Scripts

Interview Invitation Email

Dear BIOS 1120 Students,

My name is Cody Williams and I am a doctoral student with the Mallinson Institute for Science Education. Thank you for your interest in our research project aimed at the impact of different ways of introducing biological content on student learning and student attitudes towards science. Earlier this semester, 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 20-30 minutes. These interviews will be conducted here on campus in Wood Hall. You are not required to participate in the interview.

If you agree to be interviewed, you will receive 20 points extra credit towards your overall grade for the course. This research will be used for course improvement and may also help to increase what is known about the impact of how biology content is introduced on student learning and attitudes toward science. Please carefully review once more 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 at cody.t.williams@wmich.edu.

Please also send me two dates/times between April 15th and April 24th that best fit your schedule. I will send you a confirmation email with a room and time for the interview.

If you decline to be interviewed, there is an alternate assignment you may do for the same amount of extra credit. (You may not do both.) The alternative assignment must be submitted to me directly by the day of your final exam. If you would like to receive the alternate assignment please contact me by email. At the end of the term, I will inform your instructor who has completed the extra credit for grading purposes, but not whether they completed it by means of the interview or the alternative assignment.

Thank you for considering this request.

Cody T. Williams
Graduate Student
Mallinson Institute for Science Education
Western Michigan University
cody.t.williams@wmich.edu

Email for Potential Participants Not Needed in the Study

Dear ___________ ,

Thank you for your interest in participating in an interview for my research project examining pre-service teachers' conceptions of the nature of science. I have reached the maximum amount of participants needed for interviews. As a result, I do not need your participation in the study at this time. If your participation is needed at a later date, I will inform you by email. Please let me know if you have any questions. Again, thank you for your interest in the study.

Sincerely,

Cody T. Williams
Graduate Student
Mallinson Institute for Science Education
Western Michigan University
cody.t.williams@wmich.edu

APPENDIX C

Interview Protocol

1. What is your overall impression of the format of the course?
    - Did you enjoy taking a course that featured a "flipped classroom"? Please explain your answer.

2. On a regular basis your instructor made a point of raising questions about new material before you viewed the on-line lecture, read the chapter, completed the homework and completed a quiz.
    - Was this procedure of raising questions about material first before you did the outside work helpful to you? Why or why not?

3. The introduction of new material was often accompanied by the use of pre-assessment instruments, such as surveys and short writing assignments.
    - Was the process of coming up with your own answers to questions in a non-evaluative context prior to studying the chapter helpful to you? Why or why not?

4. The introduction of new material was also often accompanied by themes or stories from the history of biology.
    - Which approach do you perceive your instructor relied upon to teach your class?
    - Was it helpful to you? Why or why not?

5. In your opinion, would it be helpful or unhelpful for your instructor to use stories in class?
    - What do you see as the advantages or disadvantages of using stories in science classes?

6. Did the way your instructor introduced course content by means of broad themes and/or stories give you any insights into the practice of science?
    - Please explain your answer with an example.

7. Did the way your instructor introduced course content by means of broad themes and/or stories help you understand the content of the course?
    - Please explain your answer with an example.

8. Did the way your instructor introduced course content by means of broad themes and/or stories make you more or less comfortable learning science?
    - Please explain your answer with an example.

9. Why did your response change from the pre-SUSSI to the post-SUSSI? (If participant's response appears to be different pre- to post-survey)
    - Was there anything in particular that caused you to change your response for this item?
10. Why did your response change from the pre-genetics survey to the post-genetics survey? (If participant's response appears to be different pre- to post-survey)
    - Was there anything in particular that caused you to change your response for this item?
    - If there are any inconsistencies between the Likert response items and the open response items, students will be asked to clarify these responses

Possible prompts

- Could you expand on that?
- Could you tell me more about that?