Science & Education
https://doi.org/10.1007/s11191-019-00073-x

ARTICLE

Effects of Historical Story Telling on Student Understanding of Nature of Science

Cody Tyler Williams[1] · David Wÿss Rudge[2]

Published online: 07 September 2019

Abstract

Concepts related to the nature of science (NOS) have been considered an important part of scientific literacy as reflected in its inclusion in curriculum documents. A significant amount of science education research has focused on improving learners' understanding of NOS. One approach that has often been advocated is an *explicit and reflective* approach. Some researchers have used the history of science to provide learners with *explicit and reflective* experiences with NOS concepts. Previous research on using the history of science (HOS) in science instruction has approached HOS in many different ways and consequently has led to inconsistent findings regarding its utility for improving learning. One promising method for overcoming this inconsistency and teaching NOS with more traditional science content is using stories based in the history of science. A mixed method approach was used to determine whether and how the use of science stories influences undergraduates' understanding of NOS. Particular attention was paid to the explanations that students used for their understandings. Intervention and control groups completed the Student Understanding of Science and Scientific Inquiry (SUSSI) instrument. The intervention group was taught using two historical narratives while the control group was taught using minimal history. A subset of both groups was also interviewed regarding their SUSSI responses and their experiences in the course. Results indicated that the introduction of science stories helped participants gain a better understanding of the role of imagination and creativity in science. Participants mentioned science stories in their explanations for why they changed towards more informed views on SUSSI items related to imagination and creativity. The current study adds to a growing body of literature regarding the use of stories in the science classroom.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11191-019-00073-x) contains supplementary material, which is available to authorized users.

✉ Cody Tyler Williams
cody.t.williams@wmich.edu

David Wÿss Rudge
david.rudge@wmich.edu

Extended author information available on the last page of the article

1 Theoretical Background

The nature of science (NOS) is considered an important part of scientific literacy. The importance of NOS is reflected in its inclusion in many recent national curriculum documents both in the USA and beyond (Olson 2018). Emphasis on NOS has also led to several attempts by researchers to improve students' NOS views (Akerson et al. 2000; Kim and Irving 2010; Rudge et al. 2014). Despite significant effort to improve NOS understanding of both students and science instructors, both groups continue to struggle (Lederman & Lederman et al. 2014). One explanation 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.

Multiple researchers have advocated and provided empirical support for using an *explicit and reflective* approach to teaching NOS (Akerson et al. 2000; Rudge and Howe 2009; Williams and Rudge 2016). *Explicit* refers to planned instructional practices that allow for NOS aspects to be openly covered in class. *Reflective* refers to providing students with opportunities to consider their own learning about NOS aspects (Abd-El-Khalick and Lederman 2000). The *explicit and reflective* approach is underpinned by two theoretical frameworks, constructivism, and conceptual change theory. Constructivism posits that learners align new information with previously learned constructs through active engagement (Ausubel 1960; Matthews 1997). Conceptual change maintains that learners will seek a "best fit" between their previous understanding of a concept and new learning. This means that learners will try to match new information with previous knowledge (Appleton 1997). As a result, instructors must take care to use teaching approaches that account for learners' previous learning.

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 provides a highly contextualized approach to teaching NOS, which can introduce a human element to NOS instruction. More importantly, historical accounts of scientists are necessarily tightly bound with science content, allowing instructors to teach NOS concepts alongside of other science content (Clough 2006). HOS instruction can encourage learners to think about their own understandings of science concepts and engage in reasoning similar to past scientists (Monk and Osborne 1997; Rudge and Howe 2009).

The practical realities of the classroom emphasize the importance of melding NOS instruction with science content. There are two general approaches to teaching NOS and science content, decontextualized and contextualized. 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 (Clough 2006). Contextualized approaches are important because understandings of many NOS concepts are context-dependent. Additionally, these approaches require integration of NOS instruction with science content instruction, making it less likely that instructors will see contextualized NOS instructional approaches as wasting time (Clough 2006).

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 previous studies. 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. Stories based in the history of science are a possible solution to the difficulties with operationalizing HOS. These history-based science stories are referred to as science stories in this manuscript.

Science 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). Narrative researchers have provided empirical support for narratives being associated with improved learning in comparison to other types of communication (Dahlstrom 2014). In particular, some researchers have argued that using or communicating knowledge requires narratives. Constructing and recalling knowledge also relies upon narratives (Schank and Abelson 1995). 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). Science educators have proposed that narratives 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 and Klassen 2014).

Additionally, several scholars have advocated the use of historical narratives specifically in the context of biology education. Schwab (1958) suggested teaching biology through a series of narratives or cases from the history of science. More recently, Kampourakis (2013) discussed the importance of providing students with a more complete picture of the history of a genetics order to teach students about NOS. In particular, genetics instruction should reference a much wider range of scientists than just Mendel such as Darwin, Galton, and Nägeli. Kampourakis and McComas (2010) noted the potential for using the story of Darwin and the study of evolution for teaching students about NOS, particularly the influence of society and creativity on science. Subsequently, Kampourakis and Gripiotis (2015) developed a course focused on the development of Darwin’s theory of evolution for teaching students about the historical, cultural, and societal influences on science. The course encourages students to engage with original writings from Darwin and the literature and scientists of Darwin’s time. These writings are intended to provide students with an understanding of what shaped Darwin’s thinking and how Darwin’s work subsequently affected society.

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 and 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 and 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.

Hadzigeorgiou et al. (2012) and Tsybulsky (2018) represent two of the few primarily empirical studies published recently in science education journals. Hadzigeorgiou et al. (2012) 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 concerns related to the nature of the historical story used for the study in that it overemphasized the romantic aspects of the story. Tsybulsky (2018) found that teaching high school biology students using an approach that involved reading a historical narrative about the development of a model for DNA structure helped students improve their understanding of several NOS aspects. However, Tsybulsky (2018) approach did not make clear what their narrative entailed and also included student visits to research laboratories. As a result, it was not clear how effective the narrative was in affecting student understanding of NOS. Both the Hadzigeorgiou et al. (2012) and Tsybulsky (2018) took different approaches to narratives making comparisons between studies difficult. 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.

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 1). 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 et al. (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 are few currently available empirical studies supporting Klassen's story framework. One of studies done by Klassen himself (Klassen 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 a priori 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 did 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 that the story was effective in decreasing student misconceptions about natural selection. The

Table 1 Narrative elements from Klassen (2009)

Narrative element
Event-tokens
• Narratives are composed of a series chronologically 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 reader's 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.

current study moves beyond the Klassen (2009) study and builds off of the Fulford (2016) study.

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.

2 Method

The purpose of the current study is to build off of the recent science education literature related to stories based in the history of science and test the effect of Klassen's (2009) story framework. In particular, this study provides a systematic empirical assessment of the effects of historical narratives, based on the Klassen (2009) framework, on undergraduate students understanding of NOS concepts. There are multiple frameworks for characterizing what is meant by NOS. The current study uses the multidimensional framework (MD), described by Deng et al. (2011), which characterizes NOS as a set of agreed upon concepts such as tentativeness of scientific knowledge and the influence of creativity and society on science (Lederman et al. 2002; McComas 2004; Lederman 2007). The MD framework was chosen to maximize comparability of the findings of this study with previous studies. The majority of recent NOS studies have used this framework and most of the currently existing instruments for measuring NOS understandings are informed by the MD framework (Deng et al. 2011; Liang et al. 2008; Lederman et al. 2002).

The study compares two traditionally taught biology units in a non-major introductory biology course to the same units taught using science stories. Klassen noted ten narrative elements that should be included in a historical story intended for the classroom (Table 1). One of the science stories used in this study, adapted from B. Williams et al. (2010), is based on the work and life of Gregor Mendel. The story is used to introduce students to Mendelian genetics and basic inheritance patterns. The other story focuses on research done on industrial melanism in the twentieth century and is similarly used to introduce a unit on evolution and natural selection. A mixed methods approach was used to determine whether the use of science stories developed using Klassen's (2009) narrative elements had an effect on student understanding of NOS concepts. The quantitative portion of the study involved pre- and post-assessments for NOS and the qualitative portion included semi-structured interviews.

2.1 Research Questions

The research questions guiding this study are:

Q1. What differences in NOS understandings are revealed from pre- to post-instruction based on participants' Student Understanding of Science and Scientific Inquiry (SUSSI) Likert and open response scores in both the traditional and historical story groups?
Q2. 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?
Q3. What awareness do participants have regarding the stories and the associated narrative elements used in the course instruction?

2.2 Study Context

Participants were recruited for interviews from two sections of the same introductory non-major biology course taught by the same instructor. Recruitment was completed on the first day of class with an invitation included on the consent form for the study. Data was collected from a section of the course during both the Fall 2014 and Spring 2016 semesters. A total of 91 students completed both the pre- and post-assessments in the Fall 2014 semester and 94 completed both in Spring 2016. Both groups consisted of 70% female students. The majority of students in both semesters were white. Participants in the Spring 2016 semester were significantly older ($\overline{x} = 19.3$) than Fall 2014 ($\overline{x} = 18.6$) ($p < .05$). Spring 2016 semester participants had significantly fewer high school science courses ($\overline{x} = 3.05$) than Fall 2014

($\overline{x} = 3.48$) ($p < .05$). 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. Students recruited in the Spring 2016 semester became the intervention group. Those participants who indicated interest in participating in the interviews were contacted through email. 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. 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.

2.3 Course Instruction

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 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.

2.4 Stories Included in the Intervention

The intervention group received the same genetics instruction as described above with the exception of the addition of two science stories. Examples of how Klassen's elements were incorporated into the stories are provided from the story about the life and work of Gregor Mendel. The Mendel 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 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 that encouraged the reader to consider and reflect upon their understanding of NOS concepts. Three of the questions used during the intervention were developed by the authors of the story (B. Williams et al. 2010). An additional question, added by the current researchers, was used to replace one of the questions from the original version of the story to provide more clarity for students (Table 2). These questions were intended to focus students' attention on the particular NOS concepts that were supported by the story. 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 that students discussed in small groups. The instructor also led class discussions regarding the questions. 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 Mendel story are included below (Table 4). 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 them 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 (Tables 3 and 4). During day 2 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 2 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. 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.).
3. 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.
4. 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.

Three of the questions (1, 3, and 4) included with the story were created by B. Williams et al. (2010). Question 2 was added by the current researchers

Table 3 Questions included with moth story

Question	NOS concept
1. Scientists were fascinated that the dark form was becoming more common, and that it seemed to be occurring only in certain areas of Britain (and Continental Europe). What do you think these areas have in common? What major environmental change was happening in Britain and Continental Europe from about 1850 onwards?	Science is influenced by the society in which it is embedded.
2. What is a theory?	Theories provide a scientific basis for explanations and predictions of natural phenomena
3. When scientists have more than one theory that might account for a phenomenon, how do they choose between them?	Science is a social endeavor requiring communication and debate among a community of researchers.
4. What counts as evidence for or against a theory?	Science relies upon empirical evidence.
5. What are experiments? Are they necessary for progress in science?	Creativity is used throughout the scientific process. (Experiments are one of many tools including observations and life histories, that scientists choose among to answer questions.)

2.5 Data Collection

A mixed methods approach was used for this study that followed a fully mixed concurrent dominant status design (Leech and Onwuegbuzie 2009). 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.

2.5.1 Quantitative Data

The SUSSI instrument was used to measure NOS views across all study participants during both semesters (Liang et al. 2008). In particular, data collected from the SUSSI was used to answer research question 1 and 2. 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 (Table 5). 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. All six of the NOS aspects included the SUSSI were not included in the stories introduced to the intervention group. However, the full SUSSI was used to maintain the validity and reliability of the instrument.

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 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

Table 4 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 reader's 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 nineteenth century with brief flashes back to the 16th and seventeenth 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.

studies. Most of these studies have used instruments based on the multidimensional NOS framework such as the SUSSI and VNOS (Deng et al. 2011). Finally, the SUSSI was validated and checked for reliability with the population intended for this study (Liang et al. 2008, 2009).

2.5.2 Qualitative Data

The primary means of qualitative data collection for this study was semistructured interviews with a subsample of participants from both the intervention and control groups. Open-response

Table 5 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. (+)

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)

data was also collected from the SUSSI. Overall, 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 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 2–3. The interview questions covered two primary areas including impressions of the class format and participant responses on the SUSSI. The full interview protocol is included in the supplementary material. Interviews lasted around 30 min each, and were audio recorded. All interviews were conducted during the last 2 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.

2.6 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, semistructured interviews were conducted with a subsample 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.

2.6.1 Quantitative Data

All participant responses to the SUSSI Likert items were scored from 1 to 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 5 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).

2.6.2 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. McNemar's test was used to compare the proportion of participants that improved their open-response scores to the proportion that scored lower from the pre- to the post-SUSSI.

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 and Rossman 2011). HyperRESEARCH qualitative analysis software was used for all coding (Version 3.7.2, Researchware, Inc. 2015). The coding process began with the primary researcher developing an initial coding scheme (codebook) through multiple readings of the interview transcripts. The codebook development process outlined in this paragraph is based on processes described in Campbell et al. (2013) and Weston et al. (2001). 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.

2.6.3 Data Integration

The results for the quantitative and qualitative portions of the study were integrated after the analysis was completed for both separately (Leech and Onwuegbuzie 2009). The combined datasets were used to answer research question 2. In addition, they were used as a check 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.

3 Results

Results are summarized for the three research questions guiding this study. Research question 1 includes data regarding validity and reliability checks for the SUSSI along with the SUSSI quantitative findings. Research questions 2 and 3 are primarily focused on findings from the interviews.

3.1 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?

3.1.1 SUSSI Validity and Reliability

Validity and reliability for the SUSSI was established for the current study sample. Confirmatory factor analysis (CFA) was used to establish construct validity for the six NOS aspects on the SUSSI using a combined data set with all completed pre- and post-SUSSI instruments from both semesters included in the study($n = 454$).

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), chi-squared/*df* below 5.0 (Bollen 1989), and a comparative fit index (CFI) near 0.90 (Hu and Bentler 1999; Byrne 2013). The guidelines for acceptable values for RMSEA vary. Hu and Bentler (1995) suggested an RMSEA of 0.06

as indicative of good-fit. MacCallum and Browne (1996) suggested values of 0.01, 0.05, and 0.08 as indicative of excellent, good, and mediocre fit. CFA model fit statistics are included in Table 6. 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.

Internal reliability of the SUSSI was established for the current study using Cronbach's alpha (Table 7). 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 low. While this is consistent with the reliability values reported by the original SUSSI authors, any 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.

3.1.2 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. Figures 1 and 2 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. To assess whether there were differences between the control and intervention groups, mixed model ANOVAs were run for each construct to determine if there were any significant differences for the interaction between group (intervention, control) and time (pre-test, post-test). No significant differences were detected for the interaction term for any of the constructs indicating that there were no significant differences between control group and intervention group pre-scores or control group and intervention group post-scores.

Table 6 SUSSI confirmatory factor analysis model fit statistics

Model fit statistic	Value
RMSEA	0.05
Chi-squared/*df*	2.13
CFI	0.879

Table 7 SUSSI Cronbach's alpha values

SUSSI NOS aspect	Control		Intervention	
	Pre-test	Post-test	Pre-test	Post-test
Overall SUSSI	0.624	0.729	0.749	0.712

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 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, compared to the control group. Interview data indicated that the instruction received by participants and their experiences in the course played a role in influencing these changes. These results are discussed further in Sect. 4.2.

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

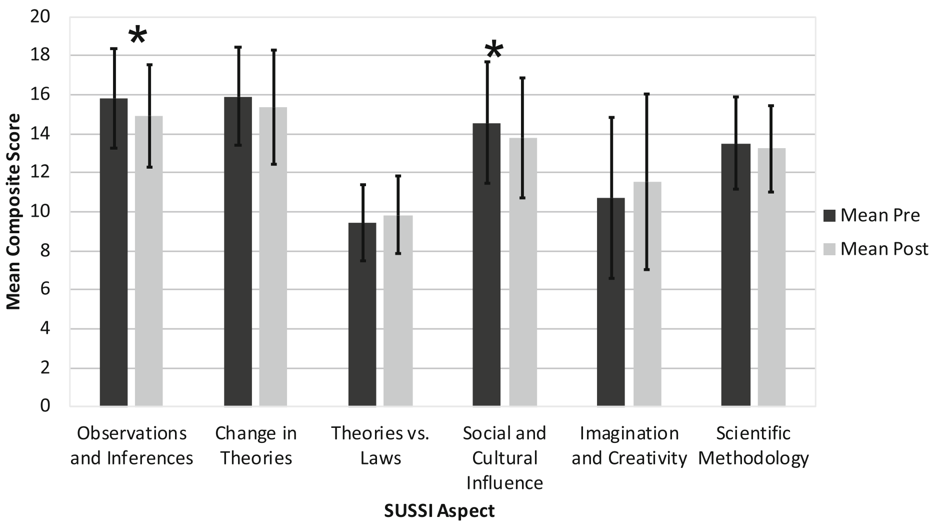

Fig. 1 SUSSI mean composite scores for control group. Error bars represent standard deviation. The asterisk (*) indicates significant decrease from pre- to post-SUSSI ($p < .05$)

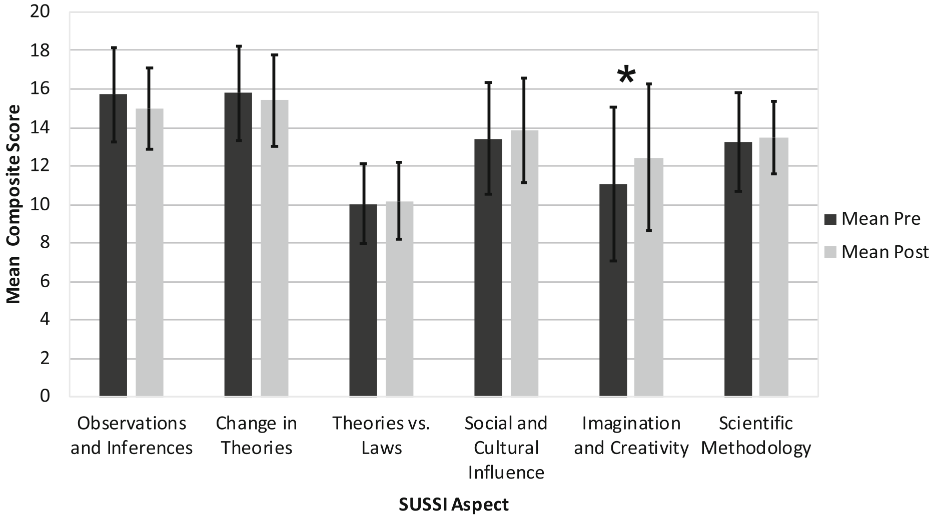

Fig. 2 SUSSI mean composite scores for the intervention group. Error bars represent standard deviation. The asterisk (*) indicates significant increase from pre- to post-instruction score ($p < .001$)

(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.

3.2 Research Question 2: 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?

3.2.1 Qualitative Results

A primary focus of the semistructured interviews was participants' responses to the pre- and post-SUSSI. Participants were provided with both of their completed surveys during the interview and were 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. In "historical explanations," participants said that the reason they changed was because of one or more historical examples from the course. In "general course" explanations, participants mentioned some aspect of the course other than historical examples. In "external influence" explanations, participants referred to something unrelated to the course that influenced their thinking. Table 8 has frequency counts

Table 8 Frequency counts of all explanation types for SUSSI Likert response changes

SUSSI NOS aspect	Control (n = 15)			Intervention (n = 12)		
	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

for each of the explanation types for both treatment groups broken down by SUSSI NOS aspect.

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 when accounting for whether the explanation was associated with a positive or negative change in SUSSI scores. Positive changes occurred 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 occurred when participants moved from more informed to less informed or their less informed views were reinforced. 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 science 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 science 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 externally influenced explanations. In the intervention group, externally influenced explanations were rarer 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 science 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 changes in the SUSSI results provide support for the science stories used in the intervention group being a positive influence on participant NOS views.

3.2.2 Connections Between SUSSI Quantitative Results and Participant Explanations

The quantitative analysis done on the SUSSI Likert data, detailed in Sect. 3.1, revealed statistically significant differences from pre- to post-instruction across both the control and intervention 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 had changed, three of these intervention group participants said that science stories from the course had influenced their thinking. One participant mentioned the Mendel story, one the moth story, and the other the science 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, where participants' SUSSI Likert scores were lower on the post-SUSSI compared to the pre-SUSSI. 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 science 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 is discussed in the results for research question 3.

3.3 Research Question 3: 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. The intent was to determine whether participants interpreted the use of history simply as recurring ideas (themes) or as some form of structured account of past events (stories), given the importance of narratives for learning and knowledge reconstruction (Dahlstrom 2014; Schank and Abelson 1995). A particular definition of stories was not expected from participants, but there was interest in whether student understandings of stories would be consistent with any of the Klassen (2009) model elements. Almost all of the intervention group participants responded that they interpreted the history as stories, whereas the majority of control group participants responded that it was themes (Table 9).

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

3.3.1 Course Historical Content Interpreted as a Story or Theme

The intervention group participants gave several reasons for 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:

Table 9 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

> 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 (F11).

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 science 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 are supported by the narrative literature (Klassen and Klassen 2014; Dahlstrom 2014).

3.3.2 Science Stories/Themes Are Useful for Learning About Science

Questions asked during the interview about whether the science stories/themes used in the course were helpful for learning about science gave valuable insights into how the stories/themes benefitted participants' 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 that scientists change their minds about how things work and the way in which 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 with 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… 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 are not particularly enjoyable. 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 science stories/themes questions largely aligned with the findings from the intervention group. The majority of control group participants (nine of fifteen) said that science stories/themes 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 science stories/themes 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 science stories/themes 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.

The three control group participants that said they had 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. 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 science stories/themes 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.

4 Limitations and Implications

Using stories based in the history of science can be problematic depending on the learning goals of the instructor. Purposeful stories from the history of science are unlikely to provide students with an adequate understanding of the history on their own. Any one historical story can introduce its own biases and privilege the perspectives of particular individuals or groups (Novick 1988; Wineburg 2001). However, this does not negate the utility of using science stories to pursue science learning goals including aspects of NOS. Instead, these concerns highlight the importance of selecting instructional approaches that are well aligned with learning goals. There were also limitations that emerged from this study that speak to potential future areas of research. Limitations related to the SUSSI instrument and the overall study design are discussed below.

4.1 Limitations Related to the SUSSI Instrument

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. SUSSI open-response data were also collected from each of the participants that completed the instrument. However, this data was less 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 does not cue participants that it is looking for both what and how, it is not surprising when both are not included.

The above limitations resulted in many responses defaulting to a score of “2.” Since the scale has difficulty differentiating between informed and less informed views, it makes it difficult to use it for meaningful analysis. 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.

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 some 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 and Millwood 2005; Deng et al. 2011).

4.2 Limitations Related to the Overall Study Design

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

The stories were also a relatively brief intervention consisting of two class periods for the Mendel story and 3 class periods for the moth story. Both stories were intended to introduce the units and grab students' attention prior to their learning course science content. A similar intervention could likely be improved in future work through 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 min). However, computer simulations tie in nicely with discussions of Mendel (Williams and Rudge 2015, 2016). This would also allow for studying how Klassen framework stories 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 science stories and how they influence their understandings.

Finally, the size of the sample for the current study, 185 participants, means there may be some generalizability to the population of students taking Biology 1120 at Western Michigan University. Particularly given that the students in both sections were similar in gender, age, previous coursework, and final grades (see Sect. 2.2). However, the sample was a non-random convenience sample meaning that there is 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).

5 Discussion and Conclusions

Improving understanding of NOS is a longstanding and uncommonly attained goal of science education. Despite significant effort to improve NOS understanding of both students and science instructors, both groups continue to struggle (Lederman & Lederman et al. 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), Williams and Rudge (2016), and Tsybulsky (2018) 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.

Defining what using the history of science in the science classroom means is important for determining its utility for teaching about NOS. One way that history of science can be defined is through stories. Klassen (2009) laid out a framework for creating 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.

Recently some science education researchers have begun advocating the use of stories based in the history of science in science instruction (Metz et al. 2007; Klassen 2009; Klassen and Klassen 2014). Klassen (2009) proposed a framework for creating science 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 science stories based on the Klassen (2009) framework for improving students NOS views or science content understanding outside of Fulford (2016).

The evidence documented in this study 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.

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 two groups was the introduction of science stories in the intervention group. These findings indicate that the science stories positively influenced the NOS views of intervention group participants related to imagination and creativity in science.

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 that 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 Sect. 3.3.1, 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.

Overall, the results revealed several insights about the dispositions of the students involved in the current study towards science stories. These results also gave some understanding of how the historical content included in both treatment groups was experienced by participants. The majority of participants across both groups were able to recognize and or articulate an understanding of stories. Many of the intervention group participants' ideas for why they thought the history presented was a story were consistent with Klassen's (2009) 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. Additionally, intervention group participants were able to discuss numerous reasons why the science 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 that they 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. The findings from this study suggest support for the use of history and in particular science stories in the science classroom as a tool for teaching students about both NOS and more traditional science content.

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. There is also a need to determine the extent to which history and stories enhance student learning. Systematic studies that compare instruction framed around historical narratives and ahistorical narratives would help to discern the importance of history and narratives per se in helping students to learn science content.

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 traditional 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.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

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.

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.

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

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education, 95*(3), 518–542.

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.

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.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Sage.

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.

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.

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.

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.

Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences, 111*(Supplement 4), 13614–13620.

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.

Fulford, J. M. (2016). Assessing the impact of historical story telling on student learning of natural selection. Retrieved from ProQuest Dissertations & Theses Global. (1846142100).

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*, 59–82.
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.
Hatcher, L., & Stepanski, E. J. (1994). *A step-by-step approach to using the SAS system for univariate and multivariate statistics*. Cary: SAS Institute.
Howard, P. J. (2000). *The owner's manual for the brain: Everyday applications from mind-brain research* (2nd ed.). Austin: Bard Press.
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: 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.
Kampourakis, K. (2013). Mendel and the path to genetics: Portraying science as a social process. *Science & Education, 22*(2), 293–324.
Kampourakis, K., & Gripiotis, C. (2015). Darwinism in context: An interdisciplinary, highly contextualized course on nature of science. *Perspectives in Science, 5*, 25–35.
Kampourakis, K., & McComas, W. F. (2010). Charles Darwin and evolution: Illustrating human aspects of science. *Science & Education, 19*(6–8), 637–654.
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.
Klassen, S. (2009). The construction and analysis of a science story: A proposed methodology. *Science & Education, 18*(3–4), 401–423.
Klassen S., & Klassen C. (2014) Science teaching with historically based stories: Theoretical and practical perspectives. In Matthews M. (Ed), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 1503–1529).
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: 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.
Lederman N.G., Bartos S.A., Lederman J.S. (2014) The development, use, and interpretation of nature of science assessments. In Matthews M. (Ed), *International Handbook of Research in History, Philosophy and Science Teaching* (pp. 971-997).
Leech, N., & Onwuegbuzie, A. (2009). A typology of mixed methods research designs. *Quality & Quantity, 43*(2), 265–275.
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.
MacCallum, R., & Browne, M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.
Marshall, C., & Rossman, G. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks: Sage Publications, Inc..
Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. New York: Routledge.
Matthews, M. (1997). Introductory comments on philosophy and constructivism in science education. *Science & Education, 6*, 5–14.
McComas, W. F. (2004). Keys to teaching the nature of science. *Science Teacher, 71*(9), 24–27.
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).
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.
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.
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.
Novick, P. (1988). *That Noble dream: The "objectivity question" and the American historical profession*. New York: Cambridge University Press.

Olson, J. K. (2018). The inclusion of the nature of science in nine recent international science education standards documents. *Science & Education, 27*(7–8), 637–660.

Onwuegbuzie, A. J., & McLean, J. E. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools,* 10(1), 71–89.

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.

Researchware, Inc (2015). HyperRESEARCH [Computer Software]. Retrieved from http://www.researchware.com. Accessed 30 Sept 2015.

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

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.

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.

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

Schank, R. C., & Abelson, R. P. (1995). Knowledge and memory: The real story. In R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 1–85). Hillsdale: Lawrence Erlbaum Associates.

Schwab, J. J. (1958). The teaching of science as inquiry. *Bulletin of the Atomic Scientists, 14*(9), 374–379.

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.

Tsybulsky, D. (2018). Comparing the impact of two science-as-inquiry methods on the NOS understanding of high-school biology students. *Science & Education, 27*(7–8), 661–683.

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.

Williams, C., & Rudge, D. (2015). Mendel and the nature of science. *The American Biology Teacher, 77*(7), 492–499.

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.

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.

Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past.* Philadelphia: Temple University Press.

Affiliations

Cody Tyler Williams [1] • David Wÿss Rudge [2]

[1] The Mallinson Institute for Science Education, Western Michigan University, 3233 Wood Hall, Kalamazoo, MI 49008-5410, USA

[2] Department of Biological Sciences & The Mallinson Institute for Science Education, Western Michigan University, 3134 Wood Hall, Kalamazoo, MI 49008-5410, USA