Assessing the Impact of a Historically Based Unit on Preservice Teachers' Views of the Nature of Science

Running head: Assessing the Impact of a Historically Based Unit

David Wÿss Rudge[1]

Department of Biological Sciences & The Mallinson Institute for Science Education
Western Michigan University

3134 Wood Hall, Kalamazoo, MI 49008-5410, U.S.A.
tel. 269-387-2779; fax 269-387-5609; david.rudge@wmich.edu

David Paul Cassidy

The Mallinson Institute for Science Education
Western Michigan University

3134 Wood Hall, Kalamazoo, MI 49008-5410, U.S.A.
tel. 269-387-2779; fax 269-387-5609; david.p.cassidy@wmich.edu

Janice Marie Fulford

The Mallinson Institute for Science Education
Western Michigan University

3134 Wood Hall, Kalamazoo, MI 49008-5410, U.S.A.
tel. 269-387-2779; fax 269-387-5609; janice.m.fulford@wmich.edu

Eric Michael Howe

Department of Education, Assumption College

159 Kennedy Hall, 500 Salisbury Street, Worcester, MA 01612
tel. 508-767-7015; emhowe@assumption.edu

[1] This manuscript was submitted to *Science & Education* on 17 Jul 2012. Please do not quote from it without permission from the authors.

Abstract: Numerous empirical studies have provided evidence of the effectiveness of an *explicit* and *reflective* approach to the learning of issues associated with the nature of science (NOS) (c.f. Abd-El-Khalick and Lederman 2000). This essay reports the results of a mixed methods association study involving 130 preservice teachers during the course of a three class unit based on history of science using this explicit and reflective approach. Within the unit the phenomenon of industrial melanism was presented as a puzzle for students to solve. These students were explicitly asked to reflect upon several NOS issues as they developed and tested their own explanations for the "mystery phenomenon". NOS views of all participants were characterized by means of surveys and follow-up interviews with a subsample of 17 participants using a modified version of the VNOS protocol (c.f. Lederman *et al.* 2002). Analysis of the survey results informed by the interview data suggests participant NOS views changed during the course of the unit. These emerging changes where positive for many nature of science views, e.g. whether scientific knowledge requires experimentation, but was neutral or negative for others, e.g. why scientists experiment. An examination of the interview data informed by our experiences with the unit provides insight into why some views changed during the course of the unit but others did not. Positive changes to some NOS issues appear to have been fostered by the use of contextualized examples. The essay concludes with a discussion of limitations, pedagogical implications, and avenues for further research. [250 words]

Key words: evolution learning; evolution teaching; explicit and reflective approach, history of science; nature of science; Views on the Nature of Science (VNOS)

1 Introduction

Science educators have identified an informed understanding of issues associated with the nature of science (NOS) as a fundamental component of science literacy for well over a century (DeBoer 1991). While consensus exists that students should learn both *of* and *about* science as indicated in national science standards documents (American Association for the Advancement of Science [AAAS] 1993, National Research Council [NRC] 1996), many studies have documented that teachers often belabor under misconceptions about NOS (e.g. Abell and Smith 1994; Akerson *et al.* 2000). As such, considerable recent attention has been paid by the science education community to the study of the effects of curricular interventions on participants' views of NOS (Deng *et al.* 2011).

The present study reports the impact of a unit based on history of science as a instructional intervention to promote NOS. Use of history of science to promote learning of NOS has been the object of considerable attention by historians and philosophers of science as well as science educators (e.g. Matthews 1994, Jenkins 1994). Numerous empirical studies have attempted to document the impact of using history of science as an instructional strategy for promoting understanding of NOS among preservice teachers (e.g. Abd-El-Khalick and Lederman 2000, Lin and Chen 2002). The purpose of the present study is to explore the impact of a historically based unit on preservice teachers understanding of NOS, with the intention of documenting not only the impact of the unit but also why it was effective.

2 Nature of Science (NOS) and the Explicit Reflective Approach

Research on the teaching and learning of NOS has been complicated by the fact that those who actually do research on NOS (i.e. historians, philosophers and sociologists of science) often disagree strongly with one another (Abd-El-Khalick *et al.* 1998). This being said, a consensus view within the science education community for some time has identified what students and teachers should learn about NOS with the epistemology of science, science as a way of knowing, and/or the values and beliefs inherent in the development of scientific knowledge (Lederman 1992, McComas *et al.* 1998, Rudolph 2000). Lederman (2007) succinctly summarizes the core tenets as follows:

> Scientific knowledge is tentative (subject to change), empirically based (based on and/or derived from observations of the natural world, and subjective (involves personal background, biases, and/or is theory-laden); necessarily involves human inference, imagination, creativity (involves the invention of explanations); and is socially and culturally embedded. Two additional important aspects are the distinction between observations and inferences, and the functions of and relationships between science theories and laws. (p. 833) [2]

---

[2] This consensus view that appears to equate understanding of NOS with a relatively short list of declarative claims has been the object of considerable recent debate. Clough (2007) contends students would be better served if practitioners discussed these issues as open-ended questions, rather than facts be memorized. Irzik and Nola (2011) present an alternative approach based on a notion of family resemblance; Allchin (2011) provides a fundamentally different approach to assessing NOS based upon a reframing of how NOS is characterized from a list of tenets to a multidimensional perspective of the practice of science. These debates unfortunately fall outside the scope of the present essay.

Historically there have been two quite distinct approaches within the science education community to the teaching of NOS: *implicit* and *explicit* (Abd-El-Khalick and Lederman 2000). *Implicit* approaches to the teaching of NOS tacitly presume that students will develop more sophisticated understandings as a result of science instruction by means of either science process-skills instruction and/or science inquiry activities on their own (Abd-El-Khalick 1998). This implicit approach to NOS appears to embody an assumption that the process of doing science alone will result in improved NOS understandings.

Explicit (often referred to as the *explicit reflective* approach) approaches, in contrast, identify NOS understandings as cognitive instructional outcomes in their own right. The explicit component refers to treating NOS as a planned instructional activity to ensure that student attention is drawn to NOS issues, by means of instructor prompts (Khishfe and Lederman 2007). The reflective component emphasizes students must be provided with opportunities to discuss and reflect on the issues when reaching their own insights, as opposed to having the instructor didactically "tell" students how the content relates to the NOS issue (Akerson *et al.* 2000). Numerous recent empirical studies have demonstrated that the explicit reflective approach is more effective than the implicit approach (e.g. Abd-El-Khalick and Lederman 2000, Khishfe and Abd-El-Khalick 2002, Hanuscin *et al.* 2006), including studies of preservice teachers (e.g. Akerson *et al.* 2000). These findings led Lederman (2007) to conclude "conceptions of NOS are best learned through explicit, reflective instruction as opposed to implicitly through experiences with simply 'doing' science" (p. 869), but see McDonald (2010).

3 The Role of History of Science

While consensus exists regarding the importance of using an explicit, reflective approach to teaching NOS, science educators disagree regarding whether and to what extent NOS instruction should be contextualized, as, for instance, when one uses history (Clough 2006). Numerous scholars within and outside the science education community have long touted the potential role history of science can play in promoting numerous aspects of science education, including student understanding of NOS. Matthews (1994 and in numerous other articles) has long advocated an integrated approach to the use of history, rather than using it as an "add on". Teachers should integrate history of science as they teach science by means of reproducing historical experiments, role playing historical debates, or reading and interpreting original scientific papers, rather than teaching the science content in isolation with occasional asides to its historical development. He points out numerous potential advantages associated with this reconceptualization of history of science as the foundation of science instruction, including humanizing the process of science, promoting critical reasoning skills, and better preparing teachers to address student misconceptions, which may be similar to ideas held by past scientists.

Monk and Osborne (1997) point out that much of the reason why science teachers at all levels have not followed the advice of Matthews and others on this score has to do with (1) their relatively impoverished views on the nature of science, which lead them to focus on products rather than process, (2) the imperatives of the classroom (e.g.

classroom management and the need to prepare students for standardized tests), and (3) the relative dearth of teacher-friendly materials that illustrate how to incorporate history of science in the classroom. Monk and Osborne therefore propose an alternative rationale for the inclusion of history of science in the classroom, one that follows from a constructivist theory of learning (cf. Driver and Oldham 1986). Instead of asking students to set aside the sincerely held, but often erroneous beliefs they have about NOS to make sense of historical actors working in a completely different context, teachers should instead find ways to use history to engage students in rethinking their own ideas. Monk and Osborne further proposed a model for how teachers might more successfully include history of science in their classrooms, one that relies on the introduction of the views of a past scientist by the teacher as an alternative point of view.

Rudge and Howe (2009) agreed with Monk and Osborne's justification for why history of science should be an integral part of the science classroom, but rejected their proposed model as being inconsistent with its stated rationale. As written, Monk and Osborne's model still obligates students to make sense of the reasoning of historical actors in their own context. Rudge and Howe proposed an alternative, "instrumental" approach to history, in which they fully embrace Monk and Osborne's contention that teachers will only start to use history when use of history is subject to the imperatives of the classroom. In particular, Rudge and Howe (cf. Allchin 1993) emphasize instructors have a license to abandon historical accuracy whenever slavish attention to historical detail for its own sake would undermine their other learning objectives.

The intervention of the present study was developed by the first author as an example of this "instrumental" approach to the use of history (specific details are provided in Section 4.2). Development of the unit began with identification of science content and NOS learning objectives. He wanted students to learn how scientists study microevolutionary phenomena and also more general NOS issues, such as how scientists test alternative explanations by means of observations and experiments. The first author recognized the potential of a particular episode in the history of biology, namely the history of research on industrial melanism, for these objectives. He knew, for instance, that many misconceptions students have about microevolutionary phenomena were at one time the object of serious scientific research on industrial melanism. The unit accordingly begins, in common with Monk and Osborne's model, with an elicitation phase in which students are asked to make sense of a set of observations: an odd correlation between areas where dark moths are becoming more common and those that have been the subject of large scale air pollution associated with the industrial revolution. Students are asked to come up with their own (often mistaken) ideas for why this trend (the mystery phenomenon) is taking place, and rather than rejecting them as false, the instructor validates them by drawing attention to the fact that a past scientist once thought along the same lines. With this background in place, students are explicitly asked to reflect on what theories are in science and how they are evaluated. Students are then encouraged to consider how they might test between alternative explanations for the mystery phenomena, and again because their authentically held ideas are entirely predictable, the instructor can use the sharing of student ideas to motivate a discussion of simplified summaries of the results of similar investigations that were actually pursued by past scientists. Students are then explicitly asked to reflect on what experiments are in science and how they are evaluated. The first author's goal throughout is to engage students with reference to the ideas they bring into his classroom, not the teaching of history of science for its own sake. This being said, he concludes the unit by debriefing students with "the rest of the story", discussing how and why what they did together departs from the messy details of the history of science. The first author's repeated experience has been that once students are familiar with the basic story they become receptive to learning the complicated details that lie behind any fact in science and the NOS issues we would like them to critically reflect upon.

Howe and Rudge (2005) reported the results of a study on the impact of an eight-class unit that used this same instrumental approach to the use of history (the history of research on sickle-cell anemia) to teach NOS. The study was conducted on a population of 81 preservice teachers at a large Midwestern university. The impact of the unit was assessed by means of a slightly modified version of the VNOS-C (Views on the Nature of Science) instrument (c.f. Lederman *et al.* 2002) that was administered pre and post intervention, follow-up pre and post interviews, and journal assignments. Their results, reported in terms of frequencies, suggested participant NOS views improved with reference to five specific issues. Rudge *et al.* (2007) reported the results of a pilot study involving 19 students of the instructional unit that is the object of the present study. They cautiously concluded on the basis of their results that the unit did affect some change on a set of targeted NOS issues. The present study used a modified version of the VNOS protocol (described in detail in Section 4.3 below) on a larger population to address two questions:

(1) Were there any changes in participants' conceptions about targeted nature of science issues associated with the intervention? and,
(2) If so, how do interviews with a subset of participants and our previous experiences with the intervention inform our interpretation of the results?

4 Method

The study was conducted using a modified version of the VNOS protocol (c.f. Abd-El-Khalick and Lederman 2000, Lederman *et al.* 2002). An abbreviated version of the *VNOS-C* (Views on the Nature of Science Version C) questionnaire (shortened owing to the fact that several questions on the original were not germane to the present intervention) was administered pre and post intervention. Interviews were conducted with a subset of the students post instruction only in view of the brevity of the unit. The impact of the unit was assessed using a mixed methods approach. Student responses to survey questions were categorized and ranked by means of an emergent coding and ranking schema, with interview data being used to clarify our interpretation of student responses. The resulting categorical and ordinal data inferred from this qualitative set were analyzed using non-parametric statistics both with respect to individual questions and the unit as a whole. To address the second question we used a qualitative approach. We reexamined interview transcripts for insights into what students perceived as the reasons for why their views might have changed and interpreted these in light of the first author's experiences teaching the unit for nine years.

4.1 PARTICIPANTS AND CONTEXT OF THE STUDY

The study was initiated under the auspices of a Human Subjects Institutional Review Board at a large Midwestern university during the Fall 2007 and Spring 2008 terms. The potential pool of participants was 221. The actual number of students who agreed to participate in the study and completed both pre- and post- surveys was 130, and among them 17 agreed to be interviewed. 93% of the participants for the study as a whole self-identified as Caucasian, 85% were female. Participants ranged in age from 18-30 years old; the mean age was 21. 17% of the participants indicated they had previously taken at least one philosophy course. 130 (100%) of the participants attended the first day of the intervention, 111 (85%) attended the second, and 100 (77%) attended the last [8 (6%) missed both]. The final grade of the total populous taking part had mean of 2.74 (on a 4.0 scale) with a standard deviation of 0.85.[3]

The intervention took place in the context of an introductory biology course, one of six science content courses taken by future elementary school teachers. Each of these courses features a guided open-ended problem solving environment in which students are encouraged to take charge of their own learning. The course in question has four units devoted to taxonomy, anatomy and physiology, ecology and evolution. At the time of the study the course was taught in a lecture lab format, with students meeting once a week for a 2 hour 'lecture' session and twice a week for a 1½ hour small section (24 student) lab.[4] The lecture section was taught by the first author during both terms. A total of 11 lab sections (6 during Fall 2007; 5 during Spring 2008) were taught. One lab section was taught by the first author during the Fall 2007 and Spring 2008 terms. The remaining nine lab sections were taught by multiple graduate student instructors, each of whom was responsible for either one or two lab sections in a given term. The instructional sequence of interest was taught in lab over three successive sessions as part of a final unit devoted to evolutionary biology. All of the instructors of lab sections had previously taught the unit of interest at least twice.

4.2 INTERVENTION

The three day instructional sequence of interest in the present study is based upon the history of research on industrial melanism, presented to students as "The Mystery Phenomenon" through a series of short background lectures, guided discussions and individual and group work. A brief review of this intervention highlighting use of the *explicit reflective* approach to the teaching of NOS (c.f. Abd-El-Khalick and Lederman 2000) is provided below, the lesson plans themselves are available here: sci-ed.org/documents/Rudge%20**Lessons**.pdf.

---

[3] A Two-sided Welch Two Sample t-test revealed males in the study population were significantly older than females [$t(19) = 2.34$, $p = 0.\ 03$]; but there was no significant difference between them with regard to achievement as measured by final grade [$t(27) = 1.29$, $p = 0.206$]. The bins for races other than Caucasian were too small (<<30) for any meaningful statistical tests to be performed.

[4] Lab sections were taught by inquiry (i.e. laboratory instructors taught primarily by means of carefully worded questions aimed at facilitating student driven discussions of and about biological topics and the process of science). Lecture sessions were devoted to practicing example problems of the sort that would appear on exams, with students attempting to solve problems on their own and with the help of the person seated next to them before the class as a whole discussed their answers.

Class 1

During the first lab session, students are introduced to a mystery first discovered with reference to *Biston betularia*, a common moth known throughout Britain and Continental Europe. The point of this discussion is to elicit without judgment student views (including misconceptions) about how this change might have occurred and to provide a basis for a general discussion of theories that follows in the next class. Students learn that when the moth was first described by the naturalist Moses Harris (1986 [1766]) it was characterized by a pale, speckled appearance, with no indication that any other form of the moth existed. A photograph of the moth as it rests on a lichened-covered tree trunk draws attention to the apparent adaptive value of its pale, speckled appearance in effectively camouflaging it against its background. The instructor then shares how the discovery of a rare dark form of this moth ca. 1848 near Manchester, England, prompted naturalists to search for more examples. Students are provided with survey maps of Britain that document ever increasing additional sightings of the dark form over the course of the next one hundred years. Attention is drawn to the fact that these sightings are not uniformly spread throughout Britain--the spread of the dark form appears to be limited to certain areas. Students are also informed that it is not merely that the range of the dark form has increased, but also that in some places, the dark form has gone from being completely unknown to very common. Students are asked to consider what else was also going on in Britain (and Continental Europe) between 1850 and 1950. Invariably at least one student mentions recognizes that this time period coincides with the Industrial Revolution, a time when the environment near manufacturing centers was dramatically changing as a result of the first large scale air pollution associated with the burning of coal, coke and oil. Students are provided with three images (two landscape portraits and a photo) of roughly the same countryside near Manchester over time (1730, 1860 and 1954). The instructor then returns attention to a map documenting relatively frequencies of dark and light forms throughout Britain, and asks students whether they notice a pattern. At this point, even with only a minimal familiarity with the geography of Britain, students notice that the dark form is most common in the vicinity of large cities, such as London, Manchester and Edinburgh, which the instructor confirms were indeed centers of manufacturing during this time period. The instructor then asks students to explicitly discuss not only why they think the dark form is becoming more common in the vicinity of these manufacturing districts, but also the reasons that suggest their ideas are plausible and worthy of further consideration. Our experience teaching the class has been that students invariably gravitate to explaining the mystery phenomenon in one of three ways. Some correctly recognize it might be explicable in terms of natural selection, but often have difficulty explaining it. Others belabor under what the instructor would recognize as based upon a misconception, either that the moths have the ability to individually change color when they need to (Lamarckian inheritance) or that it is the result of a mutation directly caused by the ingestion of toxins (air pollution) by caterpillars of the moth.

Class 2

The instructor begins the second lab class by reviewing what students have learned about the mystery phenomena so far, and the three different explanations students came up with in the previous class. The point of this class is to validate each of the explanations the students have come up with by drawing attention to how each was actually proposed by a scientist who worked on we are calling "the mystery phenomenon" during the twentieth century. Attention is drawn to the fact that in the case of the scientists, their hypotheses were developed in the context of theoretical frameworks, specifically Darwin's Theory of Natural Selection, Lamarck's Theory of the Inheritance of Acquired Characteristics and De Vries' Mutation Theory. After discussing how scientists (E.B. Ford, Nicholas Cook, and James Heslop-Harrison respectively) used these theories to account for the mystery, the instructor asks students to *explicitly* and *reflectively* (c.f. Abd-El-Khalick and Lederman 2000) discuss what theories are in general, using the following prompt:

> So far we have considered three different theories that might account for the Mystery Phenomenon. Let's take a step back. What is a theory in general?

The goal of this explicit reflective discussion is to encourage students to share their sincerely felt ideas regarding what theories are in general without evaluating or judging their ideas. This is done in part by asking students to share theories they have encountered in the context of other science courses, and then pressing them to consider what it is about their example that makes it a theory. Once the group appears to reach some consensus on what theories are, the instructor then prompts them to consider whether and how scientists choose amongst alternative theories using the following prompt:

> What do scientists do when they have more than one theory for the same phenomena? Is this a matter of everyone simply being entitled to their own opinion? Or are there ways to choose among alternatives?

Our students often recognize that scientists conduct experiments and other forms of investigation to test amongst alternatives. We explicitly ask them to consider whether scientists have still other ways to choose among alternatives, in the hope one or more may recognize that considerations of plausibility, consistency with other more accepted theories, etc. might also play a role. The class concludes by having students devise ideas for how each of the three proposed hypotheses for the mystery phenomenon might be tested by means of observation and experiment.

Class 3

The third and final class begins with a review of what students have discussed so far regarding the mystery phenomenon. Our goal in this class is to have students discuss their ideas for how each of the three proposed explanations might be tested with reference to the results of similar tests actually conducted by past scientists. Students are explicitly invited to consider simplified summaries of results of two or three investigations conducted by past scientists from the perspective of both advocates and critics of each proposed explanation. Attention to the important role that observations and experiments play in testing hypotheses provides a natural opportunity for the instructor to lead an *explicit reflective* discussion on what experiments are in science using the following prompt:

> So far we have been using the term "experiment" in a fast and loose way to describe how scientists test alternative explanations of the same phenomena. What is an experiment in general?

The goal of this explicit reflective discussion is to encourage students to share their sincerely felt ideas regarding what experiments are in general without evaluating or judging their ideas. This is done in part by asking students to share experiments they have encountered in the context of other science courses, and then pressing them to consider what it is about their example that makes it an experiment. Students sometimes offer astronomical observations as examples of experiments, which provides the instructor with the opportunity to draw attention to a restricted use of the term "experiment" in biology. (In the life sciences, "experiment" is used specifically to refer to systematic study in which a system is perturbed with reference to an independent variable and the effect of the perturbation is observed with reference to a dependent variable.) Once the group appears to reach some consensus on what experiments are, the instructor then prompts them to consider whether experiments are always necessary:

> Are experiments always necessary for scientific progress to be made?

Once more, the focus of the discussion is on student generated examples. Students are encouraged to consider examples in historical sciences, such as geology and evolution, where experimentation is not always possible.

The class concludes by having students view a film (*Evolution in Progress*) that appears to conclusively demonstrate that the phenomenon should be understood in terms of natural selection (Kettlewell 1961). The instructor then reveals ‘the rest of the story’, additional details that draw attention to the fact that the mystery phenomenon is much more complicated than textbooks would have us believe. A concluding discussion asks participants (all of whom are future elementary school teachers) to consider how they will help their students recognize how very misleading textbook accounts can be with reference to both the process and nature of science.[5]

4.3 PROCEDURE

The present study is based on a modified version of the VNOS questionnaire (see **Appendix**) and its associated protocol, which has been widely used in the literature to study NOS. The former was initially developed and refined in a series of papers by Norman Lederman and his associates (Lederman and O’Malley 1990, Lederman *et al.* 1998, Lederman *et al.* 2002). The latter has its origins in these papers, but is perhaps most explicitly first discussed in Abd-El-Khalick and Lederman (2000). We first briefly summarize the standard protocol and the motivations that led us to depart from it. In Section 6.1 we discuss the impact of these departures on our interpretation of results.

The VNOS has been typically used in the literature to study entire classes, and has often involved a comparison of two or more classes with distinct pedagogies with reference to NOS, one (or more) of which is referred to as a treatment (e.g. Abd-El-Khalick and Lederman 2000, Lin and Chen 2002, Kim and Irving 2010, Yacoubian and BouJaoude 2010). According to the suggested protocol, the instrument should be used for preassessment purposes during a single class without time limits. Student responses are

[5] Only one lecture session took place during the course of this three lab sequence. Students practiced problems that required them to distinguish evidence for natural selection from evidence for common descent. They also practiced explaining microevolutionary phenomena in terms of natural selection and did some concept mapping.

next analyzed in order to develop profiles for each student. A subset of students who participated in the preassessment should then be interviewed to ensure they have interpreted the questions as intended, and researchers are interpreting their responses as they intended. Later, after the intervention, the same VNOS survey is used once more for postassessment. These responses are similarly analyzed to develop profiles for each student. Again, a subset of students who participated in the postassessment are invited to be interviewed. The interviews are conducted similarly, only this time the interviewers ask students to also compare their pre- and post- responses. This provides the investigators with an opportunity to ask students if their views changed, and if so, why. Surveys are also provided to an outside consultant, who independently develops profiles for each student. By comparing these, the investigators are able to assess "inter-rater reliability", i.e. the extent to which their interpretations of student responses to questions are objective and not tinged by personal bias. Ultimately the goal of the analysis is to compare pre- and post- responses so as to assess whether students views have changed along a continuum from more naïve to more informed views.

As noted below, our study required us to depart from the standard protocol in several respects. First, we used the protocol to study the effect of a short term intervention, not an entire class. This is because we were interested in the effect of a single unit based upon history of science in the context of a class characterized by a multiplicity of other pedagogies. Second, we had a graduate student researcher mix pre- and post- responses to survey questions prior to any of them being coded and ranked so as to avoid having knowledge that a response was pre or post bias our interpretation. Third, the brevity of our study led us to conduct only one set of interviews, post instruction. Fourth, while we coded student responses to individual questions in light of their responses to other questions, our analysis did not involve the creation of student profiles. Fifth and finally, as noted in Section 5 below, our analysis of results includes a study of how individual student responses changed pre to post, not only the entire population in aggregate.

Open-ended Surveys

A list of the questions used before and after the instructional sequence of interest is provided as Appendix (space between items deleted). The survey included six separate questions, each question addressing a distinct NOS issue that has been emphasized in recent reform documents (AAAS 1993, NRC 1996). Question 1 was developed and validated in the context of a previous similar research project (Howe 2004). Questions 2-6 were taken from the Views of Nature of Science version C (VNOS-C), discussed in detail in Lederman *et al.* (2002). (The design of the VNOS survey and the method for its validation are discussed in detail in Lederman and O'Malley (1990).) Four of the questions had multiple parts, and one of the questions proved to be inherently ambiguous; as such, they were ultimately analyzed as twelve distinct items. Nine of these items represent issues that the Mystery Phenomenon Unit was intended to address; the remaining three items (Questions 1b, 2a and 2b) were included to retain validity of the survey instrument.

All students filled out the pre- and post- surveys as a requirement for the course. The graduate student researcher administered surveys in ten lab sections for which she was not the instructor of record in the absence of the first author. (The first author administered surveys in the eleventh lab section taught by this graduate student in her absence and subsequently gave them to her in a sealed envelope for processing.) Pre-surveys were administered at the start of the first class of the intervention; post-surveys were administered at the start of the next class after the intervention. Participants took an average of about 20 minutes to complete the survey each time.

Semi-structured Interviews
Twenty candidates were chosen at random by the graduate student researcher at the conclusion of the unit from the 130 students who completed both the pre- and post-surveys. Each was offered a small financial reimbursement ($20) in order to ensure that at least thirteen would take part. A total of 17 participants (13%) agreed to be interviewed. The interviews took place within two weeks of the conclusion of the unit and typically lasted between 30 minutes and an hour. All interviews were audio-taped and transcribed for analysis. Interviews were conducted post instruction only owing to the brevity of the unit (3 lab classes held over the course of two weeks). During the interviews the researcher asked each student to reread both the survey questions aloud and their responses. The researcher asked follow up questions aimed at clarifying whether the student understood the original question and also the specific wording of the students' answers. Students were then invited to paraphrase their responses and, when change occurred, discuss what led them to change their answers.

None of these students was interviewed by his/her laboratory instructor. The first author (who interviewed a single student) was trained to conduct interviews during a similar previous research project by the fourth author (Howe 2004). The first author trained the graduate student researcher, who interviewed the remaining 16, by having her read several relevant articles in the literature (e.g. Abd-El-Khalick and Lederman 2000; Palmquist and Finley 1997; Ginsberg 1997), study copies of student written responses to pre- and post- surveys, and mock interviews.

4.4 DATA ANALYSIS

Data was analyzed in a manner similar to established protocols in similar research into student conceptions of NOS (Abd-El-Khalick 2001, Abd-El-Khalick *et al.* 1998; Akerson *et al.* 2000, Howe 2004, Palmquist and Finley 1997). The first author read through all the pre- and post- surveys to identify emergent themes amongst student responses to each of the questions considered in isolation. For each question he was able to identify 4-10 mutually exclusive themes (codes), which provided a basis for the coding of student responses. Each response was considered once more with reference to the student's responses to other questions in the hope this would clarify the student's intended answer to the question of interest. The emergent codes were then ranked as representing more or less sophisticated views on each NOS issue with reference to standards documents (AAAS 1993, NRC 1996).

The reliability of this coding and ranking of data was later assessed by having the third author independently code (and rank) all responses to survey questions using the first author's ten page coding and ranking scheme (260 surveys total). There was substantial inter-rater agreement between the coders as measured by Cohen's Kappa, $\kappa = 0.65$ (Landis and Koch 1977). (This contrasts with previous studies in which two or more individuals develop a code together by means of consensus scoring and assess interrater reliability as a measure of consistency between coders on a sample of remaining surveys drawn at random (e.g. Lederman *et al.* 2002). We were, in essence, testing the application of the code by an independent coder.)

To analyze the impact of the intervention on each question, a Stuart-Maxwell test for marginal homogeneity was used by the second author on a matrix of the number (frequency) of students for each potential pair of codes for that question (Stuart 1955, Maxwell 1970). This established whether the result was significant. To assess the net impact, he took the total number of pairs representing improvement and subtracted the total number of pairs representing backsliding.

Demographic data was analyzed to establish that the interviewed students were representative of the study population as a whole. A statistical comparison of the subpopulation who were interviewed (n = 17) with those that were not (n = 113) reveals that they did not differ significantly with regard to race, gender, age, previous philosophy courses, or achievement. The race of those interviewed did not differ significantly from those we were not [$\chi^2(3) = 5.5$, $p = 0.14$]. The gender of those interviewed was found to be very close, but not significantly different from those we were not [$\chi^2(1) = 3.0$, $p = 0.08$]. The age of the interviewed traditional students (M = 20.5, SD = 1.3) did not differ significantly from those we were not (M = 20.5, SD = 1.5), $t(13) = 0.17$, $p = 0.87$. Those interviewed did not differ significantly from those we were not with regard to the question of whether they had ever taken a philosophy course [$\chi^2(1) = 0.032$, $p = 0.86$]. The achievement of the interviewed students, as measured by final grade, (M = 3.0, SD = 0.9) did not differ significantly from those we were not (M = 2.7, SD = 0.8), $t(18) = 1.17$, $p = 0.25$.

5 Results

As noted above, this study was intended to address two questions: (1) Were there any changes in participants' conceptions about targeted nature of science issues associated with the intervention? and (2) If so, how do interviews with a subset of participants and our previous experiences with the intervention inform our interpretation of the results? Sections 5.1 and 5.2 share results that bear on the first research question. Section 5.3 shares those results (primarily interview data) that bear directly on the second research question.

5.1 IMPACT OF THE INTERVENTION WITH RESPECT TO INDIVIDUAL QUESTIONS[6]

A summary analysis of the interpreted data collected from the surveys is provided in **Table 1**.[7] For each question it summarizes change in the relative frequencies of how

---

[6] In the text that follows, results of the present study will be compared with those of the fourth author's previous study (Howe 2004) as reported in Howe and Rudge (2005). The results of Abd-El-Khalick's (1998, 2001) previous study were not reported in such a way to allow a question by question comparison with the results of the present study.

[7] Interview data was used primarily to assess whether participants interpreted the survey questions as intended and whether written responses were being interpreted as the students intended. For the most part

responses were coded (and ranked) pre and post instruction in the population as a whole. (An elaborated version of **Table 1**, which also summarizes changes in ranking (improvement, no change, or backsliding) with reference to frequencies of paired responses (pre- and post-) for each participant is provided in the **Electronic Appendix**). A summary analysis of the impact of the intervention on each question, indicating that nine of the questions were coded significantly differently, is provided in **Table 2**. [8]

[INSERT TABLE 1]

[INSERT TABLE 2]

Question 1a-c invited participants to share their understandings of what theories are, how they are created, and provide an example of when they have used a theory. As indicated in Table 1, Question 1a, most student responses prior to instruction were coded as representing relative naïve views about the nature of theories, generally identified as an unproven claim, hypothesis or guess. Only a relatively small percentage held a more sophisticated view of theories as explanatory frameworks that provide a basis for prediction. As noted in Table 2, there was a significant difference in the distribution of codes for Questions 1a and 1c pre and post instruction: the intervention had a net positive impact (initial - final) on how these items were coded. We note further that 17 students specifically drew upon examples from the Mystery Phenomenon Unit in response to Question 1c. These results are comparable to those found in Howe's previous 2004 study, which documented a similar increase in understanding of the explanatory role of theories among student responses post instruction (from 50% (n= 42) to 70% (n = 57) [N =81]).

Question 2 a-b followed up on the first by asking students to consider whether theories ever change, and if so, why. A naïve conception often held by students is that scientific knowledge once discovered never changes, it is either wholly accepted or wholly abandoned (Cotham and Smith 1981). A more sophisticated understanding includes recognition that not only do theories change over time (as the result of new findings or technology), but that such changes may result from the reinterpretation of existing data. Prior to instruction, most of the student responses to Question 2a indicated that the student recognized that theories do change, most identifying this as a consequence of new evidence. Only a small fraction recognized theories might change as a result of the reinterpretation of data. We found no significant differences in the distribution of codes for Question 2a pre and post intervention. The intervention appears to have had a small net effect on how responses to Question 2b (the request for an example) were coded--the fact that it was nevertheless significant suggests the intervention may have reinforced whatever beliefs were already present. This being said,

---

good agreement was found between student written responses and views shared during the interviews, but some discrepancies (discussed in Section 6 below) were also found.

[8] The codes for Question 6b included one code (3, vague reference to an example from the Mystery Phenomenon Unit) that was ultimately not identified by the first author in any of the pre- or post-responses. This code was initially introduced in parallel with other items (1c, 2b and 4b) against the theoretical possibility one or more of the references to examples to the Mystery Phenomenon would be judged vague, rather than reflective. The first author chose to retain this unused code in the coding and ranking schema against the possibility the independent coder would identify an example from the Mystery Phenomenon Unit as only a vague reference among responses to Question 6b.

17 students specifically drew upon examples from the Mystery Phenomenon Unit in response to this question. These results contrast with those found in Howe's previous study, which documented a decline in student post instructional responses denying that theories can change (from 12% (n= 10) to 2% (n = 2) [N =81]) and a slight increase in student recognition that change could be a consequence of the reinterpretation of data (from 4% (n= 3) to 7% (n = 9) [N =81]).

Question 3 asked student to define what they think an experiment is. Written responses to this question in the context of the survey (and oral responses during the interviews) led us to conclude that students found the question inherently ambiguous. It was accordingly analyzed twice, first in terms of what the response said regarding what the point of an experiment is (3a), and second in terms of what the conduct of an experiment involves (3b). With regard to the question of what the point of an experiment is, a naïve conception identifies experiments as tests. A more sophisticated understanding draws attention to the direct role experiments play in testing hypotheses, the indirect bearing of results of such tests on theories. Prior to instruction, about a third of respondents identified experiments as tests of hypotheses and a sixth identified experiments as tests of theories. We found a significant difference in the distribution of codes for Question 3a pre and post instruction: the intervention appears to have had a net negative impact. With regard to the question of what does the conduct of an experiment involve (Question 3b), a naïve view identifies experiments with the collection of evidence. A more sophisticated understanding includes recognition that experiments in the context of biology involve perturbing the system involved and comparing results with a similar system not so perturbed (the experimental and control arms of the study, respectively). Prior to instruction about a third of the responses indicated recognition that experiments in biology are distinct from other forms of data collection in that they involve manipulation. We found a significant difference in the distribution of codes for Question 3b pre and post instruction: the intervention had a net positive impact (initial - final) on how this item was coded.[9] One student provided an unsolicited example from the Mystery Phenomenon Unit in response to this question:

> In a controlled experiment (not natural environment) that involve running types of tests or trials in support of trying to figure something out. Example is the test (experiment) of Harrison's theory by giving certain leaves to moths. (Student 15, post-survey)

Question 4 invited students to consider whether experiments were essential for the development of scientific knowledge. A scientifically literate person recognizes that scientific knowledge can change for other reasons (e.g. observations and the reinterpretation of existing data). Prior to instruction, most of the participants claimed experiments were required for the development of scientific knowledge, but many responses identified experiments with data collection. Only a relatively small proportion denied experiments were necessary. We found a significant difference in the distribution of codes for Question 4a pre and post instruction: the intervention had a net positive impact (initial - final) on how this item was coded. These results are comparable to those found in Howe's previous study, which documented a similar decline in student post instructional responses denying that scientific knowledge can develop in the absence of

---

[9] Question 3 was not included in surveys used in Howe (2004).

experiments (from 52% (n= 22) to 17% (n = 8) [N =42][10]). His study also demonstrated an increase in student recognition that observational evidence alone can lead to the development of scientific knowledge (from 5% (n= 2) to 43% (n = 21) [N =42]). We also found a significant difference in the distribution of codes for Question 4b pre and post instruction: the intervention appears to have had a net negative impact. This being said, six students cited examples from the Mystery Phenomenon Unit. This contrasts with the findings of Howe's previous study, in which an increase in the number of responses citing examples was observed (17% (n=7) vs. 36% (n=15) [N=42]).

Question 5 asked students to account for how it is possible that two groups of scientists looking at the same data could reach different conclusions. A naïve conception is one that attributes this to limits in available data, i.e. we don't know enough yet to decide between them. A more sophisticated conception includes recognition that the same data might be interpreted differently by virtue of differences in scientists' theoretical and experimental frameworks. Prior to instruction, a majority recognized the possibility that scientists might interpret the data differently. We also found a significant difference in the distribution of codes for Question 5a pre and post instruction: the intervention had a net positive impact (initial - final) on how this item was coded. These results are comparable to those found in Howe's previous study, which documented a similar increase in student post instructional responses mentioning the possibility that scientists could interpret the same data differently (from 54% (n= 44) to 64% (n = 59) [N =81]). Four students provided unsolicited examples from the Mystery Phenomenon Unit in response to this question:

> Even if both groups have the same set of data, they are not going to have the same ideas of how the dinosaurs became extinct. For example, the mystery phenomenon everyone in the class had different ideas of how the butterfly was dark colored. (Student 46, post-survey)

> Two different conclusions are possible because there can be different ways to interpret the same set of data. For example, a decrease in the number of dark moths could be due to lack of camouflage in a non-polluted forest, but it could also mean that they have a second predator in the environment. One set of data doesn't necessarily yield one answer. (Student 55, post-survey)

> It is possible because neither group of scientists have discredited the other. And it is possible to have more than one theory for event it does not make either of them wrong example the three theories for the light moths and the theories for the silver box. (Student 95, post-survey)

> It's just like when we did the Mystery Phenomenon, we came up with 3 theories that fit the information and some we didn't look at as well. People's background & beliefs and education will affect a theory. Is one better than the other, it depends on the tests and experiments done to provide evidence to get to a conclusion to that answer. (Student 126, post-survey)

---

[10] The total population of students asked this question in Howe's (2004) study was 42 rather than 81 because of a change in its wording between semesters when the survey was administered.

Finally, Question 6 invited students to consider whether imagination and creativity play any role in scientific investigations. A naïve conception may acknowledge that imagination plays some role, but often portrays this as an aberration to be avoided. A more sophisticated understanding draws attention to the role of imagination and creativity in all stages of the investigation, from coming up with the design to figuring out how to collect data and interpret it. Prior to instruction, most recognized that imagination and creativity play some role in scientific investigations, but most limited it to planning and design only. We found no significant differences in the distribution of codes for Question 6a pre and post instruction. With regard to Question 6b, only about a third of the pre- and post- responses included valid examples of creativity that gave some indication of reflectivity. Eleven students offered examples taken from the Mystery Phenomenon Unit. [11] Most responses to this question were difficult to interpret because many students either failed to share anything or gave such an abstract answer that it fell short of providing a specific example:

> Scientist [sic] have to use creativity and imagination during investigations. By planning and designing theres creativity in the design and inner workings of the expirement [sic]. Once data is collected imagination must be relied on to determine what the numbers could possibly conclude. Other reasons for creativity is [sic] sometimes scientists cannot see anything they are hypothesizing about such things in space or the past. Therefore they rely on their own imagination to fill in the blanks. (Student 96, post-survey)

5.2 IMPACT OF THE INTERVENTION AS A WHOLE

The impact of the intervention as a whole was assessed by the second author using a Wilcoxon Signed-rank Test (c.f. Wilcoxon 1945). For each question, a rank was assigned to each student's pre-intervention response. These ranks were then normalized and summed for all 12 test items so as to assign a test rank. This was done again for the student's post responses and then all student ranks pre- and post- were compared using the one-sided Wilcoxon Signed-rank Test, with continuity correction (H_0: true location shift (pre-post) ≤ 0 vs. H_1: true location shift (pre-post) > 0). The post- student ranks were ranked significantly better (higher ranking) $p < 0.01$. The effect-size (-0.50) was measured by the difference of location calculated using the Hodges-Lehmann estimator. This difference was the shift of location from pre to post on ten point ranking, one being the best ten being the worst, whose results ranged from four to ten both pre and post.

5.3 IMPACT OF THE INTERVENTION WITH RESPECT TO INDIVIDUAL QUESTIONS

The foregoing analysis of the aggregate survey results suggests that while the intervention had a significant, positive net impact on the distribution of codes assigned to student responses to questions associated with a select set of NOS issues, the unit was more effective with regard to some issues than others. It was found that though the written responses to the VNOS questionnaire elucidated how student conceptions changed due to the intervention, it provided few insights into what aspects of the intervention led to these changes--namely when an example from the Mystery

[11] Question 6(a & b) was not included in surveys used in Howe (2004).

Phenomenon Unit was mentioned in a response. Interviews were more insightful in this regard because during the interviews the researcher had an opportunity to ask students not only if their views had changed, but also why they believed the change occurred. Within this subsection we reexamine the results of post instructional interviews with 17 participants for insights into what it was about the unit that either aided or detracted from student understanding of targeted NOS issues.

It should be emphasized that the interviews were conducted prior to the coding of survey responses, and as such, there was no way the researchers could have known antecedently whether any of the volunteers had exhibited the desired change. During the interviews each participant was explicitly asked to compare his/her pre- and post-instruction responses. For each question, interviewees were specifically asked whether they noticed a difference between their pre- and post- responses, and if so, what led them to change their views. As noted below, this introduced a potential interpretive problem in that students might be more likely to downplay their pre-survey responses. The summary below will focus on responses to several of the questions for which the intervention had a significant impact on how responses were coded.

Question 1. What is a theory?

As noted in **Table 2**, the intervention had a significant, positive net impact on the distribution of codes assigned to student understandings of what theories are. Nine of the seventeen students who agreed to be interviewed exhibited some improvement in their understanding of what theories are, as indicated by a comparison of how their pre- and post- instructional responses to survey questions were coded and ranked (see examples in the first row of **Table 3**). Of these nine, five chose examples from the Mystery Phenomenon Unit in their written responses to Question 1c. During the interviews, participants were nevertheless vague in identifying what it was about the unit that led to a change in their views, e.g.:

> I[nterviewer]: Mmm-hmm. Good. Alrighty, and then on – have you ever used a theory… before… in your own… experiences?
> S[tudent]: Yeah.
> I: Okay. Can you give me an example?
> S: Well, the one I gave was about the one in class where we had, uh, the… the… the *Betularia* [?] or whatever…
> I: The (unintelligible)? Mmm-hmm.
> S: Yeah and um, we had to come up with three theories on why they were changing col— why the, the, uh, the colors, you know, were changing. And we came up with, you know, three different things and it was kinda… I mean, we created our own theories and then compared them to what scientists came up with.
> I: How did you do that? How did you come up with your…
> S: Um, we used, like, the idea of natural selection and, uh, a couple – I don't remember what else we did but we, um, used that idea and then explained how – why that would be occurring.
> I: Okay. (Student 117 interview, 2->1)[12]

[12] For this quotation and others in this section, we identify in parenthesis: 1) the number randomly assigned to the participant, 2) the source of the quotation, and 3) how the participant's response to the question was scored pre- to post. In general, lower numbers represent codes that were ranked as more sophisticated.

Student responses during the interviews and the teaching experience of the first author suggest the significant, positive net impact of the intervention on the distribution of codes assigned to this question was likely due to the explicit reflective discussion of theories that occurred during the second class. During this discussion participants were specifically asked what theories are and what was it about Darwin’s Theory of Natural Selection, Lamarck’s Theory of the Inheritance of Acquired Characteristics and De Vries’ Mutation Theory that makes them theories. Students were also asked to share examples of theories they have encountered in other science classes. We believe asking students to reflect on what multiple examples of theories drawn from different contexts have in common is what led to this change, as were our instructors' attempts to tease out what the distinction is between a theory is in science and how the term "theory" is used in other contexts. Seventeen students specifically drew upon examples from the Mystery Phenomenon Unit in their responses to Question 1c.

Question 2. Explain why you think theories do (or do not) change?
The intervention was unsuccessful in affecting the distribution of codes assigned to student responses to Question 1b "How are theories created?" and 2a "Explain why you think that scientific theories do (or do not) change." This was expected, as the instructional unit did not focus on examples or discussion of how theories are created or change, and this question was only asked so as to comply at the highest levels of efficacy with the VNOS-C. There was also a significant change in the distribution of codes, corresponding to a net decline in rank after the intervention when students were asked to provide written examples for Questions 2b, 4b and 6b.[13] This often contrasted with examples students were able to provide during interviews. This finding suggests students may have run out of time (or were unwilling to take the time) to provide examples in the anonymous context in which they completed the surveys, but were more engaged when they were interviewed.

Question 3. What is the point of an experiment? What does an experiment involve?
The intervention appears to have confused students with regard to why scientists carry out experiments: many were unclear as to whether experiments are tests of theories or hypotheses. Only three of the interviewed students exhibited some improvement. While none of the interviewed students' responses pre- to post- represented backsliding, we suspect in retrospect that way the unit was taught gave the false impression that theories are tested directly by observations and experiments.

Six of the seventeen students who agreed to be interviewed nevertheless exhibited some improvement in their understanding of the distinction between experiments and other types of empirical inquiry used by scientists. During the interviews all of the respondents whose views changed in the favored direction agreed their views changed; two identified the cause of the change to be in class activities:

> S: Ok. Um, I um, I remembered learning more about the experiment, the control group and experimental group in here. Like you have a control arm and an experimental arm of

[13] The decline in rank for Question 6b was not determined to be statistically significant, because, as discussed in Section 3.1 above, Question 6b could not be analyzed using the Stuart-Maxwell test for marginal homogeneity.

the, of it. So I think I, that kind of added to my, because we learned that a long time ago, but refreshing it while we were in class made me think about those again." (Student 39 interview, 4->1)

I: Ok. And why do you think you changed your answer a little bit to include the control and experimental group and it wasn't mentioned in the pre-survey. Do you recall why that is?
S: I think this probably, like my post-survey, probably has more to do with in class, just the way we were talking about things, and just the different things we've gone over. So it might just be I had more information or…
I: Ok, I see…
S: That may be why…to be completely honest I don't know why I changed my answer, but I, I'm also positive we talked about this, my post-survey, in class. (Student 71 interview, 2->1)

Again, student responses during the interviews and the teaching experience of the first author suggest the significant, positive net impact of the intervention on the distribution of codes assigned to Question 3b was due to an explicit reflective discussion of experiments that occurred during the third class. During this discussion participants were specifically asked to consider what a variety of experiments previously discussed as tests of various explanations of the Mystery Phenomenon, have in common. Students were also asked to share examples of experiments they have encountered in other science classes. As noted in our discussion of Question 1 above, we believe asking students to reflect on what multiple examples drawn from different contexts have in common is what led to this change, as were the attempts of instructors to elicit the distinction between experiments and other types of evidence, such as observations.

Question 4. Does the development of scientific knowledge REQUIRE experiments?
Four of the seventeen students who agreed to be interviewed exhibited some improvement in their understanding that scientific knowledge can develop by means other than experiments, as indicated by a comparison of how their pre- and post-instructional responses were coded and ranked (see examples in the eighth row of Table 3). Two of the students were particularly forthcoming in why their views had changed, and, as with the previous pilot study (Rudge *et al.* 2007), both students drew upon examples from the earth sciences, e.g.:

I: Sure. Alrighty. So… I'm going to ask this question once more. "Does the development of scientific knowledge require experiments?"
S: Uh… I don't – I want to say yes because that's, I mean, that, just for me, that's how I would want to do it. You know, like, that I would feel more confident if I actually did an experiment than if I just had a theory. I'm trying to think of a different way I could… you know, concretely feel like my, my theory was accurate if…
I: Um, okay. Um, in the – excuse me, I just have to look at the notes to see what you said – in the pre-survey, you said, "Yes. The development of scientific knowledge does require experiments. If you don't test your theories, you can only assume them to be accurate. An example is the relationship between the *Brassica rapa* (yawns) and the cabbage white butterfly. If you did not test the relationships, you could only assume that there was a relationship." And in the post, you said, "No. Someone may base their theories on observation. If you watch animals in the wild, you can come up with conclusions for their behavior without testing the accuracies." Oh –

> S: I guess, I guess what I was – I mean, I still think both of those. I guess my thing is whether an observation is a type of experiment.
> I: Good. Let's say for the moment that it isn't… Then it's – so we have two different ways to test a hypothesis. One might involve controls and experimentals [sic] and the other might involve just observations. Is it possible that the development of scientific theory could proceed with observations? Or would they have to have, use a theory?
> S: No, I think that they could with observations because you can't, I don't feel like you could completely use an experiment to test everything. Test all theories.
> I: Okay. Are there any instances where you just can't do experiments?
> S: Um… I don't know, like, something, like a theory about the world. I mean it's not like you can, you have two worlds and you can alter one. I mean maybe you can (laughs). I mean, someday that might, might happen where they could do that. But it seems like there are theories about the earth, you know, that you can't, you can't take the earth and put it in a little room and –
> I: Do you think biology's like that too?
> S: I think so.
> I: Alright.
> S: I think so. (Student 90 interview, 4->1)

The significant, positive net impact of the intervention on the distribution of codes assigned to this question appears to be due to the explicit reflective discussion of experiments that occurred during the third class. During this discussion participants were explicitly asked to consider whether experiments are always necessary for scientific progress to occur. Students were also asked to share examples of other instances in the history of science when progress is seen to have taken place in the absence of experimentation.

Question 5. How are DIFFERENT CONCLUSIONS possible from the SAME SET OF DATA?

Five of the seventeen students who agreed to be interviewed exhibited some improvement in their understanding of how scientists might reach different conclusions from the same set of data, as indicated by a comparison of how their pre and post instructional responses were coded and ranked (see examples in the tenth row of Table 3). One student mentioned the Mystery Phenomena Unit when interviewed about this question.

> I: Okay, so, number 5 says, [Reads Question 5] So, what do you think today? How would you answer that today?
> S: Well, this is a… more of a difficult one for me, an interesting one. It all goes back to… as a scientist, we're still a human and we, we're going to look at data and interpret it differently. Um, you know, two scientists can get the same data but it may spike something else in their mind of the way it's connected to the rest of the data that they're receiving. It's just like, you know, how I talked about the moths and the different theories that they gained. Alright, so we still have the same data. The data that there's this percentage of dark moths and this percentage of light moths, but why? In each, in each theory, there's the same data but the… to construct scientific knowledge is deeper than just reading data, you have to think about why this happened and the connections that occur. Um… I'm not really all that familiar with what is left after a meteorite hits the earth or what is left after a massive volcanic eruption, but it's possible that they may be very similar so two scientists may look at it and see, "Okay, well, I have this element

> and I have this element, this must be a volcanic eruption." Or, "I have this and this, this has to be a crater." There's no real way to, you can't go back and do the event over and see what happened, so there's going to be theories and, there's, there's never going to really be a way to pinpoint and see which theory is correct. (Student 112 interview, 1->1)

Again, it appears the significant, positive net impact of the intervention on the distribution of codes assigned to this question was due to the explicit reflective discussion of how different scientists might respond to the results of experiments and observations that occurred during the second class.

Question 6. Do scientists use their creativity and imagination during their investigations? The intervention was unsuccessful in affecting the distribution of codes assigned to student responses to this question. Again, this makes some sense, as the instructional unit did not include examples or discussion of the process of investigation. Students were provided with simplified summaries of the results of investigations by past scientists during the second class and asked to comment on their bearing on the theories being discussed, but were not made privy to the details of how these investigations were conducted. This is nevertheless disappointing, in that students were specifically asked to come up with a way to test each theory on their own before these results were shared. The process of coming up with ideas for how to test theories apparently did not lead to recognition that imagination and creativity play an important role in science. It is important to note that the instructors did not explicitly ask students to reflect on how they developed their ideas for potential tests of the theories being discussed. All this being said, three of the interviewed students mentioned referred to the Mystery Phenomena Unit in their post-survey responses, and two discussed it during the interview, e.g.:

> I: Okay, so let me just compare what you said pre and post. In the pre, you said, "Scientists have to use their imaginations because they would not know how to create the settings and supply what is needed to carry out an experiment." And the post, you said, "Yes, because if they don't use their imagination and creativity, they can't come up with the appropriate experiments or even a sufficient theory or way to collect the data. For example, in the moths experiments, one of the scientists added birds to an environment to create, to simulate what the moths would do if they had a predator among them. It took creativity to make the environment and imagination to think of adding the birds." Ah ha. So, the big difference between the two is that you've got this example from our mystery phenomena. And I'm just sort of curious, so, do you, do you see the, what, um, you had to do – you were just talking about your plant investigation, the fact that it had creativity involved in it. Do you suppose scientists have to do the same sort of, in other words, they have to be creative as well?
> S: Yes.
> I: Yes?
> S: Yes.
> I: Okay. Good. Alrighty and what was it about the unit, in this particular example with the moths, that lead you to believe that?
> S: Well… with this particular one, like I said, with him putting the birds in there, most people would not think to use a bird as a predator. They might use, um, another human in it…
> I: Uh-huh.
> S: …or something other than a bird in order to create a predator. Another type of creature to add in order to create a predator.

I: Sure. Okay, thank you so much. (Student 95 interview 6->1)

6 Discussion

Our study investigated the efficacy of a three class instructional unit developed with reference to the history of research on industrial melanism that used an explicit reflective approach to the teaching of NOS issues. Our investigation was aimed at addressing two research questions:

(1) Were there any changes in participants' conceptions about targeted nature of science issues associated with the intervention? and,
(2) If so, how do interviews with a subset of participants and our previous experiences with the intervention inform our interpretation of the results?

The foregoing analysis indicates that while the unit as a whole did have a significant, positive impact on the distribution of codes assigned to student responses, it was more effective on some issues than others. These findings are similar to those found in previous studies of explicit reflective approaches to the teaching of NOS (e.g. Abd-El-Khalick 2001, Khishfe and Abd-El-Khalick 2002, Khishfe and Lederman 2007). They also support a previous study on the effectiveness of such an approach in the context of a history-based unit (Howe and Rudge 2005). Further research is needed to assess how much of the effect is due to the explicit reflective approach and how much is due to our use of history. Our results, like those of preceding studies, suggest the explicit reflective approach is necessary, but further that the use of multiple examples of theories and investigations from the history of science may help students to appreciate what disparate examples have in common.

6.1 LIMITATIONS

We recognize a strength of the study, namely our innovative use of the VNOS protocol to study a single unit, is at the same time a liability. On the one hand, it allowed us to gain some insight into whether and how a single unit affects student NOS understandings. On the other hand, the short duration of the unit made it logistically difficult to use other measures. Thus a major limitation of the present study is that it uses the VNOS protocol as a single measure of student NOS views (Elby and Hammer 2001, Nagasawa 2004, Southerland *et al.* 2005). It should be noted that while the VNOS protocol is generally advocated with reference to the study of entire courses, rather than short units, our questions were specific to a single unit in the context of a course that uses a variety of teaching techniques. We intentionally chose to use it in connection with a single unit out of concern that the "noise" of other features of the course would drown out any information we might gain about this particular unit.

Our study also departed from the standard VNOS protocol by mixing student pre- and post- survey responses prior to coding. We believe this improved the analysis by removing one source of potential bias. The brevity of the unit led us for logistical reasons to only conduct post intervention interviews. Our protocol for the post intervention interviews included opportunities for interviewees to clarify their pre intervention responses and how they initially interpreted the question. As noted in Section 5.3 above, this introduced an interpretive problem in that students might be more likely to downplay their pre-survey responses. Nevertheless, we believe the primary impact of having pre-instructional interviews would have been to simply reduce the potential pool of candidates for the post interviews. Our study did not involve the tedious process of creating profiles for each student. Deng *et al.* (2010), in an authoritative review of empirical research on NOS draw attention to the fact that the VNOS protocol is an example of an approach based in a multi-dimensional framework that treats student NOS views as if these views were largely independent from one another. While Deng *et al.* criticize such studies for doing so (p. 981), their finding raises questions regarding how the creation of profiles *per se* informs the interpretation of results.

6.2 PEDAGOGICAL IMPLICATIONS

As noted above, the study gave us some specific insights into what it was about the intervention that led to changes in the distribution of codes assigned to student responses pre and post instruction. As noted above, our study focused on whether and how a unit based upon the history of research on industrial melanism might affect student views on issues associated with the nature of theories, experiments and the role of subjectivity and creativity in science.

With regard to the nature of theories, while significant positive changes in the distribution of codes were observed to occur took place with reference to the topic of what theories are, the intervention was less successful with respect to whether and how theories change over time and the ability of students to come up with examples. Our experience that students appear to be better able to grasp what theories are in the presence of multiple examples dovetails nicely with Galili's (2012) analysis of the importance of doing so when using examples from the history of science in the context of physics. Instructors of this unit need to pay greater attention to helping their students recognize how theories, laws and hypotheses differ from one another, where they come from and whether and how they change over time. Our study additionally suggests instructors should encourage students to come up with multiple examples so that they can appreciate the general features that these examples have in common.

Our intervention appeared to have mixed results when it came to helping students appreciate what experiments are and why they are used in biology. At least part of the problem is that the term "experiment" is used in biology to refer to studies that involve perturbing a system to see what happens in contrast to strictly observational studies. When students discuss experiments and how they are used in science, instructors need to ask students to explicitly consider what is being tested (the theory or a hypothesis).

Finally, the study led us to recognize the importance of explicitly asking students to reflect upon whether and how the conduct of an investigation might involve creativity. We now recognize the lost opportunity we had to raise this issue during the second class of the unit, when students are asked to come up with ways of testing three alternative explanations for the Mystery Phenomenon.

6.3 AVENUES FOR FUTURE RESEARCH

The present study suggests several fruitful areas for future research. First, more research needs to be done to establish the validity of the present findings and their generalizability to other contexts and other NOS aspects. We were able, in principle, to demonstrate how the study of a short term unit could lead to targeted improvement of that unit. Thus one avenue for future study might involve refinement of the protocol to provide more specific feedback regarding what it was about the unit that led to changes in student views. Second, recognition that some episodes in the history of science have more potential than others for illustrating particular NOS aspects suggests future research should be done on the effectiveness of entire courses using multiple historically based units.

Numerous authors have also suggested misconceptions about evolution in general and natural selection in particular may have their roots in misunderstandings about nature of science (e.g. Settlage, J. 1994; McComas, W. 1997; Farrari, M. and Chi, M. 1998; Dagher, Z. and Boujande, S. 2005). The present study focused on whether the use of history using an explicit, reflective approach to the teaching of nature of science might foster improvement in student understanding of several targeted NOS issues. A natural follow up would be to simultaneously collect data pre and post instruction on both participants' NOS views and their views on natural selection and evolution to see whether a correlation between the two exists.

7 Conclusions

The present study documents how a modified version of the VNOS protocol can be used to assess changes in student views associated with some NOS aspects even in the context of a short term historically based unit. It also provides suggestions for how this modified protocol can be used to distinguish those features of the intervention that worked from those that were less effective. This being said, further research is required to establish causal claims suggested by this association study.

Acknowledgements

This material is based on work supported by the National Science Foundation Grant No. 0202923. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors here acknowledge the assistance of Professor Joshua Naranjo, Department of Statistics, Western Michigan University with regard to the statistical analyses; Phyllis Haugabook Pennock, who administered surveys and conducted most of the interviews; Katherine Rowbotham and Brandy Ann Skjold, who wrote up transcripts from the interviews; and Charles Henderson and Heather Petcovic, who gave us comments on an earlier draft of this manuscript. Thanks also to the participants at the IHPST 2009 conference who attended our session, during which a preliminary report of our findings was shared and participants of the International History of Science in Science Education (8^{th} ICHSSE) Conference held at Maresias, Brazil.

References

Abd-El-Khalick, F. (1998). *The influence of history of science courses on student conceptions of the nature of science*. Unpublished doctoral dissertation, Oregon State University, Corvalis, OR.

Abd-El-Khalick, F. (2001). Do history of science courses influence college students' views of the nature of science? Paper presented at the *6th Annual Meeting of the International History, Philosophy and Science Teaching Group*, Denver, CO.

Abd-El-Khalick, F., Bell, R., & Lederman, N. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education, 82*(4), 417-436.

Abd-El-Khalick, F., & Lederman, N. (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.

Abell, S. K., & Smith, D. C. (1994). What is science? Preservice elementary teachers' conceptions of the nature of science. *International Journal of Science Education, 16*(4), 475-487.

Akerson, V. L., Abd-El-Khalick, F., & Lederman, N. (2000). Influence of a reflective explicit activity-based approach on elementary teacher's conceptions of the nature of science. *Journal of Research in Science Teaching, 37*(4), 295-317.

Allchin, D. (1993). Of squid hearts and William Harvey. *Science Teacher 60*(7), 26-33.

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

American Association for the Advancement of Science (1993). *Science for all Americans*. Oxford University Press, New York.

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. (2007). Teaching the nature of science to secondary and post-secondary students: Questions rather than tenets. *Pantaneto Forum 25* (Jan). URL: www.pantaneto.co.uk/issue25/clough.htm. Downloaded 29 Jun 2012.

Cotham, J. C., & Smith, E. L. (1981). Development and validation of the conceptions of scientific theories test. *Journal of Research in Science Teaching, 18*(5), 387-96.

Dagher, Z. & Boujande, S. (2005). Students' perceptions of the nature of evolutionary theory. *Science Education 89*(3), 378-391.

DeBoer, G. E. (1991). *A history of ideas in science education: Implications for practice*. Teachers College, New York.

Deng, F., Chen, D-T., Tsai, C-C., & Chai, C-S. (2011). Students' views of the nature of science: A critical review of research. *Science Education, 95*(6), 961-999.

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

Elby, A., Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education, 85*(5), 554-567.

Farrari, M. & Chi, M. (1998). The nature of naïve explanations of natural selection. *International Journal of Science Education 20*(10), 1231-1256.

Galili, I. (2012). Promotion of Cultural Content Knowledge Through the Use of history and philosophy of science. *Science & Education 21*(9):1283-1316.

Ginsberg, H. P. (1997). *Entertaining the child's mind: The clinical interview in psychological research and practice*. Harper Collins, New York.

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

Harris, M. (1986 [1766]). *The Aurelian or natural history of English insects; namely, moths and butterflies*. Newnes, Twickenham, UK.

Howe, E. M. (2004). *Using the history of research on sickle-cell anemia to affect preservice teachers' conceptions of the nature of science*. Unpublished doctoral dissertation. Western Michigan University, Kalamazoo, MI.

Howe, E. M., & Rudge, D. W. (2005). Recapitulating the history of sickle-cell anemia research: Improving students' NOS views explicitly and reflectively. *Science & Education 14*(3-5), 423-441.

Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education, 20*(7-8), 567-607.

Jenkins, E. W. (1994). HPS and school science education: Remediation or reconstruction? *International Journal of Science Education, 16*(6), 613-623.

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

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

Khishfe, R., & Lederman, N. (2007). Relation between instructional context and views of nature of science. *International Journal of Science Education, 29*(8), 939-961.

Kim, S. Y., & Irving, K. E. (2010). History of science as an instructional context: Student learning in genetics and the nature of science. *Science & Education, 19*(2), 187-215.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lederman, N. (1992). Students' and teachers' conceptions about the nature of science: A review of the research. *Journal of Research in Science Teaching, 29*(4), 331-359.

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

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., & Niess, M. (2001). Not-so-common "common" knowledge. *School Science and Mathematics, 101*(1), 1-4.

Lederman, N., & O'Malley, M. (1990). Students' perceptions of tentativeness of science: Development, use, and sources of change. *Science Education, 74*(2), 225-239.

Lederman, N., Wade, P., & Bell, R. L. (1998). Assessing understanding of the nature of science: A historical perspective. In W. F. McComas (Ed.). *The nature of science in science education*. Kluwer, Netherlands.

Lin, H., & Chen, C-C. (2002). Promoting preservice chemistry teachers understanding about the nature of science through history. *Journal of Research in Science Teaching, 39*(9), 773-792.

Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. Rutledge, New York.
Maxwell, A. E. (1970). Comparing the classification of subjects by two independent judges. *British Journal of Psychiatry, 116*, 651-655.
McComas, W. (1997). The discovery of nature of evolution by natural selection: Misconceptions and lessons from the history of science. *American Biology Teacher 59* (8), 492-500).
McComas, W. F., Almazroa, H., & Clough, M. P. (1998). The nature of science in science education. *Science & Education, 7*(6), 511-532.
McDonald, C. V. (2010). The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *Journal of Research in Science Teaching, 47*(9), 1137-1164.
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.
Nagasawa, P. (2004). Epistemological relevations in students’ science writing. Paper presented at the *National Association for Research in Science Teaching*, Vancouver, Canada.
National Research Council (1996). *National science education standards*. National Academy Press, Washington D.C.
Palmquist, B., & Finley, F. (1997). Preservice teachers’ views of the nature of science during a post baccalaureate science teaching program. *Journal of Research in Science Teaching, 34*(6), 595-615.
Rudge, D. W., Geer, U. C., & Howe, E. M. (2007). But is it effective? Assessing the impact of a historically-based unit. *Ninth International History, Philosophy & Science Teaching (IHPST) Conference,* University of Calgary, Calgary, Canada, Session 4.0.3 (http://www.ucalgary.ca/ihpst07/abstracts_thu.htm).
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.
Rudolph, J. L. (2000). Reconsidering the 'nature of science' as a curriculum component. *Journal of Curriculum Studies, 32*(3), 403-419.
Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching 31*(5), 449-457.
Southerland, S. A., Johnston, A., Sowell, S., & Settlage, J. (2005). Perhaps triangulation isn’t enough: A call for crystallization as a methodological referent in NOS research. Paper presented at the *AERA 2005 Annual Meeting* (Montreal, Canada) http://digitalcommons.uconn.edu/terc_docs/1.
Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika, 42*(3-4), 412-416.
Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80–83.
Yacoubian, H.A. & BouJaoude, S. (2010). The effect of reflective discussions following inquiry-based laboratory activities on students' views of nature of science. *Journal of Research in Science Teaching, 47*(10), 1229-1252.

Appendix: VNOS Survey

1. Often in science we hear words like "theories" to describe scientific knowledge.
    (a) What is a theory?
    (b) How are theories created?
    (c) Give an example of when you have created or used a theory?

2. After scientists have developed a scientific theory (e.g. atomic theory, theory of gravity), does the theory ever change?

    If you believe that scientific theories do change:
        (a) Explain why.
        (b) Defend your answer with examples.

    If you believe that scientific theories do <u>not</u> change:
        (a) Explain why.
        (b) Defend your answer with examples.

3. What is an experiment?

4. Does the development of scientific knowledge **require** experiments?
    If yes, explain why and give an example to defend your position.
    If no, explain why and give an example to defend your position.

5. It is believed that about 65 million years ago the dinosaurs became extinct. Of the hypotheses formulated by scientists to explain the extinction, two enjoy wide support. The first, formulated by one group of scientists, suggests that a huge meteorite hit the earth 65 million years ago and led to a series of events that caused the extinction. The second hypothesis, formulated by another group of scientists, suggests that massive and violent volcanic eruptions were responsible for the extinction. How are these **different conclusions** possible if scientists in both groups have access to and use the **same set of data** to derive their conclusions?

6. Scientists perform experiments/investigations when trying to find answers to the questions they put forth. Do scientists use their creativity and imagination during their investigations?

    If you believe <u>yes</u>, scientists do use imagination and creativity,
    (a) Explain why, indicating which stages this occurs (planning and design, data collection, after data collection).
    (b) Defend your answer with examples.

    If you believe no, scientists do not use imagination and creativity,
    (a) Explain why.
    (b) Defend your answer with examples.

Table 1. Change in the coding and ranking of student responses (N = 130)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 1a: What is a theory?					
Change in Code Pre->Post Instruction	1	Explanation	Most	28.5% (n=37)	38.5% (n=50)
	2	Claim		33.8% (n=44)	39.2% (n=51)
	3	Hypothesis/Guess		24.6% (n=32)	11.5% (n=15)
	4	Nonsensical/unclear/no answer	Least	13.1% (n=17)	10.8% (n=14)
Question 1b: How are theories created?					
Change in Code Pre->Post Instruction	1	Reflecting on prior knowledge	Most	21.5% (n=28)	14.6% (n=19)
	2	Direct result of empirical research		57.7% (n=75)	63.1% (n=82)
	3	Coming up with an idea		14.6% (n=19)	13.1% (n=17)
	4	Nonsensical/unclear/no answer	Least	6.2% (n=8)	9.2% (n=12)
Question 1c: Give an example when you have created or used a theory.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0.8% (n=1)	13.1% (n=17)
	2	Reflective use of other scientific example		42.3% (n=55)	50.8% (n=66)
	3	Vague reference to example from MPU		0% (n=0)	0.8% (n=1)
	4	Vague reference to other scientific example		14.6% (n=19)	9.2% (n=12)
	5	Invalid reference/non-scientific example	Least	26.9% (n=35)	14.6% (n=19)
	6	Nonsense/no example		15.4% (n=20)	11.5% (n=15)
Question 2a: Explain why you think that scientific theories do (or do not) change.					
Change in Code Pre->Post Instruction	1	Yes, if the theory is reinterpreted	Most	7.7% (n=10)	6.2% (n=8)
	2	Yes, as a result of new evidence		73.8% (n=96)	80.8% (n=105)
	3	Yes, everything is subject to change		10% (n=13)	5.4% (n=7)
	4	No, theories are discarded/replaced	Least	7.7% (n=10)	7.7% (n=10)
	5	Inconsistent (yes and no)		0.8% (n=1)	0% (n=0)
Question 2b: Defend your answer with examples.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0% (n=0)	13.1% (n=17)
	2	Reflective use of other scientific example		53.8% (n=70)	39.2% (n=51)
	3	Vague reference to example from MPU		0% (n=0)	1.5% (n=2)
	4	Vague reference to other scientific example		7.7% (n=10)	6.2% (n=8)
	5	Invalid reference/non-scientific example	Least	3.8% (n=5)	1.5% (n=2)
	6	Nonsense/no example		34.6% (n=45)	38.5% (n=50)

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 3a: What is an experiment? (What is the point of an experiment?)					
Change in Code Pre->Post Instruction	1	Test of a hypothesis	Most	36.1% (n=47)	23.8% (n=31)
	2	Test of a hypothesis or theory		10.0% (n=13)	10.8% (n=14)
	3	Test of a theory		16.1% (n=21)	26.2% (n=34)
	4	Test (unspecified)		14.6% (n=19)	19.2% (n=25)
	5	Nonsensical/unclear/no answer	Least	23.1% (n=30)	20.0% (n=26)
Question 3b: What is an experiment? (What does the conduct of an experiment involve?)					
Change in Code Pre->Post Instruction	1	Manipulating system and observing change	Most	28.5% (n=37)	43.1% (n=56)
	2	Collecting evidence		22.3% (n=29)	19.2% (n=25)
	3	Use of the scientific method		0.8% (n=1)	0.0% (n=0)
	4	Nonsensical/unclear/no answer	Least	48.5% (n=63)	37.7% (n=49)
Question 4a: Does the development of scientific knowledge REQUIRE experiments?					
Change in Code Pre->Post Instruction	1	No, science can develop by other means	Most	13.1% (n=17)	16.2% (n=21)
	2	No, experiments not always appropriate		0.8% (n=1)	9.2% (n=12)
	3	Yes, only experiments can prove		19.2% (n=25)	23.1% (n=30)
	4	Yes, data collection is necessary		47.7% (n=62)	39.2% (n=51)
	5	Yes (nonsense/no reason)		0.8% (n=1)	3.8% (n=5)
	6	Unclear OR no answer	Least	18.5% (n=24)	8.5% (n=11)
Question 4b: Defend your position with examples.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0.0% (n=0)	4.6% (n=6)
	2	Reflective use of other scientific example		35.4% (n=46)	20.0% (n=26)
	3	Vague reference to example from MPU		0.0% (n=0)	0.8% (n=1)
	4	Vague reference to other scientific example		21.5% (n=28)	16.9% (n=22)
	5	Invalid reference/non-scientific example		0.8% (n=1)	1.5% (n=2)
	6	Nonsense/no example	Least	42.3% (n=55)	56.2% (n=73)
Question 5: How are DIFFERENT CONCLUSIONS possible from SAME SET OF DATA?					
Change in Code Pre->Post Instruction	1	Differences in interpretation	Most	53.8% (n=70)	60.8% (n=79)
	2	Data is insufficient to decide between them		32.3% (n=42)	32.3% (n=42)
	3	Both might be true		6.1% (n=8)	2.3% (n=3)
	4	Nonsensical/unclear/no answer	Least	7.7% (n=10)	4.6% (n=6)

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 6a: Do scientists use their creativity and imagination during their investigations?					
Change in Code Pre->Post Instruction	1	Yes, all stages; explanation	Most	20.8% (n=27)	20.8% (n=27)
	2	Yes, all stages; no explanation		2.3% (n=3)	6.2% (n=8)
	3	Yes, P&D, data collection; explanation		3.1% (n=4)	4.6% (n=6)
	4	Yes, P&D, data collection; no explanation		2.3% (n=3)	0.8% (n=1)
	5	Yes, P&D only; explanation		46.9% (n=61)	42.3% (n=55)
	6	Yes, P&D only; no explanation		4.6% (n=6)	3.1% (n=4)
	7	Yes, no specific stage; explanation		6.2% (n=8)	7.7% (n=10)
	8	Yes, no specific stage; no explanation		3.8% (n=5)	6.2% (n=8)
	9	No		8.5% (n=11)	6.2% (n=8)
	10	Nonsensical/unclear/no answer	Least	1.5% (n=2)	2.3% (n=3)
Question 6b: Defend your answer with examples.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0.0% (n=0)	8.5% (n=11)
	2	Reflective use of other scientific example		33.8% (n=44)	24.6% (n=32)
	3	Vague reference to example from MPU		0.0% (n=0)	0.0% (n=0)
	4	Vague reference to other scientific example		2.3% (n=3)	2.3% (n=3)
	5	Invalid reference/non-scientific example		1.5% (n=2)	0.8% (n=1)
	6	Nonsense/no example	Least	62.3% (n=81)	63.8% (n=83)

Table 2. Impact of intervention on the ranking of paired student responses (pre and post) to each question

Question	χ^2	Degrees of Freedom	*p*-value	Net effect
Question 1a: What is a theory?	12.12	3	0.007	significant, 15
Question 1b: How are theories created?	3.93	3	0.269	insignificant, -10
Question 1c: Give an example when you have created or used a theory.	42.42	5	<.001	significant, 25
Question 2a: Explain why you think that scientific theories do (or do not) change.	4.22	4	0.377	insignificant, 5
Question 2b: Defend your answer with examples.	10.75	4	0.030	significant, -3
Question 3a: What is an experiment? (What is the point of an experiment?)	18.93	5	0.002	significant, -8
Question 3b: What is an experiment? (What does the conduct of an experiment involve?)	51.81	3	<.001	significant, 23
Question 4a: Does the development of scientific knowledge REQUIRE experiments?	15.26	5	0.009	significant, 27
Question 4b: Defend your position with examples.	16.78	5	0.005	significant, -23
Question 5: How are DIFFERENT CONCLUSIONS possible from SAME SET OF DATA?	10.71	3	0.013	significant, 13
Question 6a: Do scientists use their creativity and imagination during their investigations?	9.68	9	0.377	insignificant, -2
Question 6b: Defend your answer with examples.	14.82	4	0.005	significant, -2

Table 3. Representative quotations from interviewed students

Nature of Science Issue	*Less Sophisticated NOS Views*	*More Sophisticated NOS Views*
1a. What is a theory?	"A theory is something people believe to be true" (Student 129, pre-survey, 3)	"A theory is an explanation of why or how something happens for which there is evidence that suggest that the theory is correct but no concrete evidence that proves it." (Student 129, post-survey, 1)
1b. How are theories created?	"Theories are created by any one or by anything with a valuable reason yet it has to make sense" (Student 92, pre-survey, 3)	"Theories are created from someone wanting to find out the answer to something they have been noticing or an observation" (Student 92, post-survey, 1)
1c. Give an example.	"We are using and studying math theories to look at areas of shapes" (Student 111, pre-survey, 5)	"Use theory of evolution to talk about changes in organisms" (Student 111, post-survey, 2)
2a. Do theories change? Why?	"Scientific theories do change that is why they are theories and not laws" (Student 112, pre-survey, 5)	"Yes, theories constantly change. A scientist will explain what they have found, then another scientist will prove that wrong or to be inaccurate" (Student 112, post-survey, 2)
2b. Give an example.	"The theory that the heart has a open circulatory system. It then was changed to a closed circulatory system." (Student 112, pre-survey, 4)	"The theory that betularia [peppered moths] are changing from light to dark because of mutation. This would mean <u>every</u> betularia born would go through the mutation." (Student 117, post survey, 1)
3a. What's the point of an experiment?	"An experiment is a way to test a theory or a hypothesis" (Student 16, pre-survey, 2)	"An experiment is a test of a hypothesis to attept [sic] at proving or supporting a theory" (Student 16, post-survey, 1)
3b. What does an experiment involve?	"An experiment is when you test something out. You may have a theory or hypothesis and you can do an experiment to provide evidence for or against your original beliefs" (Student 129, pre-survey, 4)	"An experiment is when you are testing something, and observing to see what the outcomes are. Generally you have a control group and an experimental group" (Student 129, post-survey, 1)
4a. Does the development of scientific knowledge require experiments?	"Yes, because in order to prove some kind of scientific knowledge you have to test it to see if it is in fact a theory." (Student 16, pre-survey, 4)	"Not always. Like the example you just gave us in class with scientist looking at if there was another planet outside of Uranus. The scientist made calculations and found Neptune. No experiment required." (Student 16, post-survey, 4)

4b. Defend your answer with an example.	"[Y]ou can also gain scientific knowledge through observation." (Student 129, pre-survey)	"Scientific knowledge can be gained through observation as well. There are not experiments to figure out what the composition of the earth is. Scientist just observe and take note of what they find" (Student 129, post-survey)
5. How are different conclusions possible from the same data?	"These different conclusions are possible because they can only use the evidence that is left behind and have no way of knowing which is correct. The two ideas may be closely related." (Student 90, pre-survey, 2)	"Scientist may read the data differently and therefore come to different conclusions about what caused the extinctions." (Student 90, post-survey, 1)
6a. Do scientists use creativity and imagination in their investigations?	"Scientist have to use their imaginations because they didn't they would not know how to create the settings and supply what is needed to carry out the experiment." (Student 95, pre-survey, 5)	"Yes because if they don't use their imagination and creativity they can't come up with the appropriate experiments or even a sufficient theory or way to collect the data." (Student 95, post-survey, 3)
6b. Defend your answer with an example.	"It's like cooking: you may prepare a baked chicken the same way everytime, but the next you might say, hey I have some fruit and I'm going to add that and see what happens" (Student 92, pre-survey, 5)	"For example, we learned that LaMarck [Harrison] cheated on his theory and therefore used his imagination and creativity to alter results" (Student 92, post-survey, 1)

ELECTRONIC APPENDIX

Table 1. Change in the coding and ranking of student responses (N = 130)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 1a: What is a theory?					
Change in Code Pre->Post Instruction	1	Explanation	Most	28.5% (n=37)	38.5% (n=50)
	2	Claim		33.8% (n=44)	39.2% (n=51)
	3	Hypothesis/Guess		24.6% (n=32)	11.5% (n=15)
	4	Nonsensical/unclear/no answer	Least	13.1% (n=17)	10.8% (n=14)

Change in Rank Pre->Post Instruction						
	Improvement 33.8% (n=44)					
	2->1	3->1	4->1	3->2	4->2	4->3
	10.0% (n=13)	9.2% (n=12)	3.1% (n=4)	4.6% (n=6)	5.4% (n=7)	1.5% (n=2)
	No Change 43.8% (n=57)					
	1->1	2->2	3->3	4->4		
	16.2% (n=21)	18.5% (n=24)	6.2% (n=8)	3.1% (n=4)		
	Backsliding 22.3% (n=29)					
	1->2	1->3	1->4	2->3	2->4	3->4
	10.8% (n=14)	1.5% (n=2)	0.0% (n=0)	2.3% (n=3)	3.1% (n=4)	4.6% (n=6)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 1b: How are theories created?					
Change in Code Pre->Post Instruction	1	Reflecting on prior knowledge	Most	21.5% (n=28)	14.6% (n=19)
	2	Direct result of empirical research		57.7% (n=75)	63.1% (n=82)
	3	Coming up with an idea		14.6% (n=19)	13.1% (n=17)
	4	Nonsensical/unclear/no answer	Least	6.2% (n=8)	9.2% (n=12)

Change in Rank Pre->Post Instruction						
	Improvement 20.0% (n=26)					
	2->1	3->1	4->1	3->2	4->2	4->3
	6.2%(n=8)	2.3% (n=3)	0.0%(n=0)	6.9%(n=9)	3.8% (n=5)	0.8% (n=1)
	No Change 52.3% (n=68)					
	1->1	2->2	3->3	4->4		
	6.2%(n=8)	40.8%(n=53)	3.8% (n=5)	1.5% (n=2)		
	Backsliding 27.7% (n=36)					
	1->2	1->3	1->4	2->3	2->4	3->4
	11.5%(n=15)	3.1%(n=4)	0.8% (n=1)	5.4% (n=7)	5.4% (n=7)	1.5% (n=2)

Table 1. Change in the coding and ranking of student responses (con’t)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 1c: Give an example when you have created or used a theory.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0.8% (n=1)	13.1% (n=17)
	2	Reflective use of other scientific example		42.3% (n=55)	50.8% (n=66)
	3	Vague reference to example from MPU		0% (n=0)	0.8% (n=1)
	4	Vague reference to other scientific example		14.6% (n=19)	9.2% (n=12)
	5	Invalid reference/non-scientific example	Least	26.9% (n=35)	14.6% (n=19)
	6	Nonsense/no example		15.4% (n=20)	11.5% (n=15)

Change in Rank Pre->Post Instruction								
	Improvement 31.5% (n=41)							
	4->1	5->1	6->1	4->2	5->2	6->2	5->4	6->4
	1.5% (n=2)	2.3% (n=3)	1.5% (n=2)	4.6%(n=6)	11.5%(n=15)	7.7%(n=10)	0.8% (n=1)	1.5% (n=2)
	No Change 56.2% (n=73)							
	1->1	2->2	4->4	5->5	6->6	2->1	5->6	6->5
	0.8% (n=1)	26.9% (n=35)	4.6%(n=6)	10.0%(n=13)	3.8%(n=5)	6.9%(n=9)	2.3% (n=3)	0.8% (n=1)
	Backsliding 12.3% (n=16)							
	2->3	2->4	2->5	2->6	4->5	4->6		
	0.8% (n=1)	2.3% (n=3)	3.1%(n=4)	2.3% (n=3)	0.8% (n=1)	3.1%(n=4)		

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 2a: Explain why you think that scientific theories do (or do not) change.					
Change in Code Pre->Post Instruction	1	Yes, if the theory is reinterpreted	Most	7.7% (n=10)	6.2% (n=8)
	2	Yes, as a result of new evidence		73.8% (n=96)	80.8% (n=105)
	3	Yes, everything is subject to change		10% (n=13)	5.4% (n=7)
	4	No, theories are discarded/replaced	Least	7.7% (n=10)	7.7% (n=10)
	5	Inconsistent (yes and no)		0.8% (n=1)	0% (n=0)

Change in Rank Pre->Post Instruction					
	Improvement 18.5% (n=24)				
	2->1	3->1	4->1	3->2	5->2
	3.8%(n=5)	1.5% (n=2)	0.8% (n=1)	5.4%(n=7)	6.9%(n=9)
	No Change 66.9% (n=87)				
	2->2	3->3	5->5	5->3	
	63.1%(n=82)	0.8% (n=1)	1.5% (n=2)	1.5% (n=2)	
	Backsliding 14.6% (n=19)				
	1->2	1->3	1->5	2->3	2->5
	5.4%(n=7)	0.8% (n=1)	1.5% (n=2)	4.6%(n=6)	2.3(n=3)

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 2b: Defend your answer with examples.					
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0% (n=0)	13.1% (n=17)
	2	Reflective use of other scientific example		53.8% (n=70)	39.2% (n=51)
	3	Vague reference to example from MPU		0% (n=0)	1.5% (n=2)
	4	Vague reference to other scientific example		7.7% (n=10)	6.2% (n=8)
	5	Invalid reference/non-scientific example	Least	3.8% (n=5)	1.5% (n=2)
	6	Nonsense/no example		34.6% (n=45)	38.5% (n=50)

Change in Rank Pre->Post Instruction								
	Improvement 18.5% (n=24)							
	4->1	5->1	6->1	4->2	5->2	6->2	6->3	6->4
	0.8% (n=1)	0.8% (n=1)	4.6% (n=6)	2.3% (n=3)	0.8% (n=1)	7.7% (n=10)	0.8% (n=1)	0.8% (n=1)
	No Change 60.8% (n=79)							
	2->2	4->4	6->6	2->1	5->6			
	28.5% (n=37)	2.3% (n=3)	20.8% (n=27)	6.9% (n=9)	2.3% (n=3)			
	Backsliding 20.8% (n=27)							
	2->3	2->4	2->5	2->6	4->6			
	0.8% (n=1)	3.1% (n=4)	1.5% (n=2)	13.1% (n=17)	2.3% (n=3)			

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 3a: What is an experiment? (What is the point of an experiment?)					
Change in Code Pre->Post Instruction	1	Test of a hypothesis	Most	36.1% (n=47)	23.8% (n=31)
	2	Test of a hypothesis or theory		10.0% (n=13)	10.8% (n=14)
	3	Test of a theory		16.1% (n=21)	26.2% (n=34)
	4	Test (unspecified)		14.6% (n=19)	19.2% (n=25)
	5	Nonsensical/unclear/no answer	Least	23.1% (n=30)	20.0% (n=26)

Change in Rank Pre->Post Instruction							
	Improvement 23.8% (n=31)						
	2->1	3->1	4->1	5->1	5->2	5->3	5->4
	0.8% (n=1)	3.1% (n=4)	4.6% (n=6)	3.1% (n=4)	3.1% (n=4)	3.8% (n=5)	5.4% (n=7)
	No Change 46.2% (n=60)						
	1->1	2->2	3->3	4->4	5->5	2->3	2->4
	12.3% (n=16)	0.8% (n=1)	8.5% (n=11)	3.8% (n=5)	7.7% (n=10)	3.8% (n=5)	1.5% (n=2)
	3->2	3->4	4->2	4->3			
	1.5% (n=2)	2.3% (n=3)	0.8% (n=1)	3.1%(n=4)			
	Backsliding 30.0% (n=39)						
	1->2	1->3	1->4	1->5	2->5	3->5	4->5
	4.6% (n=6)	6.9% (n=9)	6.2% (n=8)	6.2% (n=8)	3.1% (n=4)	0.8% (n=1)	2.3% (n=3)

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 3b: What is an experiment? (What does the conduct of an experiment involve?)					
Change in Code Pre->Post Instruction	1	Manipulating system and observing change	Most	28.5% (n=37)	43.1% (n=56)
	2	Collecting evidence		22.3% (n=29)	19.2% (n=25)
	3	Use of the scientific method		0.8% (n=1)	0.0% (n=0)
	4	Nonsensical/unclear/no answer	Least	48.5% (n=63)	37.7% (n=49)

Change in Rank Pre->Post Instruction (Question 3b)

Improvement 36.2% (n=47)			
2->1	3->1	4->1	4->2
10.0% (n=13)	0.8% (n=1)	14.6% (n=19)	10.8% (n=14)
No Change 45.4% (n=59)			
1->1	2->2	4->4	
17.7% (n=23)	4.6% (n=6)	23.1% (n=30)	
Backsliding 18.5% (n=24)			
1->2	1->4	2->4	
3.8% (n=5)	6.9% (n=9)	7.7% (n=10)	

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 4a: Does the development of scientific knowledge REQUIRE experiments?					
Change in Code Pre->Post Instruction	1	No, science can develop by other means	Most	13.1% (n=17)	16.2% (n=21)
	2	No, experiments not always appropriate		0.8% (n=1)	9.2% (n=12)
	3	Yes, only experiments can prove		19.2% (n=25)	23.1% (n=30)
	4	Yes, data collection is necessary		47.7% (n=62)	39.2% (n=51)
	5	Yes (nonsense/no reason)		0.8% (n=1)	3.8% (n=5)
	6	Unclear OR no answer	Least	18.5% (n=24)	8.5% (n=11)

Change in Rank Pre->Post Instruction (Question 4a)

Improvement 36.2% (n=47)									
3->1	4->1	6->1	3->2	4->2	6->2	4->3	5->3	6->3	6->4
2.3% (n=3)	4.6% (n=6)	2.3% (n=3)	3.1% (n=4)	3.1% (n=4)	2.3% (n=3)	9.2% (n=12)	0.8% (n=1)	2.3% (n=3)	6.2% (n=8)

No Change 48.5% (n=63)				
1->1	3->3	4->4	6->6	1->2
6.9% (n=9)	10.0% (n=13)	25.4% (n=33)	5.4% (n=7)	0.8% (n=1)

Backsliding 15.4% (n=20)							
1->3	1->4	1->6	2->4	3->4	3->5	4->5	5->6
0.8% (n=1)	3.8% (n=5)	0.8% (n=1)	0.8% (n=1)	3.1% (n=4)	0.8% (n=1)	3.1% (n=4)	2.3% (n=3)

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*		
Question 4b: Defend your position with examples.							
Change in Code Pre->Post Instruction	1	Reflective use of example from MPU	Most	0.0% (n=0)	4.6% (n=6)		
	2	Reflective use of other scientific example		35.4% (n=46)	20.0% (n=26)		
	3	Vague reference to example from MPU		0.0% (n=0)	0.8% (n=1)		
	4	Vague reference to other scientific example		21.5% (n=28)	16.9% (n=22)		
	5	Invalid reference/non-scientific example		0.8% (n=1)	1.5% (n=2)		
	6	Nonsense/no example	Least	42.3% (n=55)	56.2% (n=73)		
Change in Rank Pre->Post Instruction	**Improvement 15.4% (n=20)**						
	4->1	6->1	4->2	6->2	6->4		
	0.8% (n=1)	1.5% (n=2)	3.1% (n=4)	6.9% (n=9)	3.1% (n=4)		
	No Change 51.5% (n=67)						
	2->2	4->4	6->6	2->1	4->3	5->6	6->5
	10.0% (n=13)	6.9% (n=9)	30.0% (n=39)	2.3% (n=3)	0.8% (n=1)	0.8% (n=1)	0.8% (n=1)
	Backsliding 33.1% (n=43)						
	2->4	2->6	4->5	4->6			
	6.9% (n=9)	16.2% (n=21)	0.8% (n=1)	9.2% (n=12)			
Question 5: How are DIFFERENT CONCLUSIONS possible from SAME SET OF DATA?							
Change in Code Pre->Post Instruction	1	Differences in interpretation	Most	53.8% (n=70)	60.8% (n=79)		
	2	Data is insufficient to decide between them		32.3% (n=42)	32.3% (n=42)		
	3	Both might be true		6.1% (n=8)	2.3% (n=3)		
	4	Nonsensical/unclear/no answer	Least	7.7% (n=10)	4.6% (n=6)		
Change in Rank Pre->Post Instruction	**Improvement 23.1% (n=30)**						
	2->1	3->1	4->1	4->2	4->3		
	13.8% (n=18)	3.8% (n=5)	1.5% (n=2)	2.3% (n=3)	1.5% (n=2)		
	No Change 63.8% (n=83)						
	1->1	2->2	3->3	4->4	3->2		
	41.5% (n=54)	16.9% (n=22)	0.8% (n=1)	2.3% (n=3)	2.3% (n=3)		
	Backsliding 13.1% (n=17)						
	1->2	1->4	2->4				
	10.8% (n=14)	1.5% (n=2)	0.8% (n=1)				

Table 1. Change in the coding and ranking of student responses (con't)

	Code	*Description*	*Rank*	*Pre Instruction*	*Post Instruction*
Question 6a: Do scientists use their creativity and imagination during their investigations?					
Change in Code Pre->Post Instruction	1	Yes, all stages; explanation	Most	20.8% (n=27)	20.8% (n=27)
	2	Yes, all stages; no explanation		2.3% (n=3)	6.2% (n=8)
	3	Yes, P&D, data collection; explanation		3.1% (n=4)	4.6% (n=6)
	4	Yes, P&D, data collection; no explanation		2.3% (n=3)	0.8% (n=1)
	5	Yes, P&D only; explanation		46.9% (n=61)	42.3% (n=55)
	6	Yes, P&D only; no explanation		4.6% (n=6)	3.1% (n=4)
	7	Yes, no specific stage; explanation		6.2% (n=8)	7.7% (n=10)
	8	Yes, no specific stage; no explanation		3.8% (n=5)	6.2% (n=8)
	9	No		8.5% (n=11)	6.2% (n=8)
	10	Nonsensical/unclear/no answer	Least	1.5% (n=2)	2.3% (n=3)

Change in Rank Pre->Post Instruction

Improvement 26.2% (n=34)

2->1	5->1	7->1	9->1	4->2	5->2	8->2	5->3	9->3
0.8% (n=1)	6.9% (n=9)	1.5% (n=2)	0.8% (n=1)	0.8% (n=1)	0.8% (n=1)	1.5% (n=2)	3.1% (n=4)	0.8% (n=1)
5->4	6->5	7->5	8->5	9->5	9->6	10->7	10->8	
0.8% (n=1)	1.5% (n=2)	1.5% (n=2)	0.8% (n=1)	2.3% (n=3)	0.8% (n=1)	0.8% (n=1)	0.8% (n=1)	

No Change 46.2% (n=60)

1->1	5->5	6->6	7->7	8->8	9->9
10.8% (n=14)	28.5% (n=37)	0.8% (n=1)	1.5% (n=2)	0.8% (n=1)	3.8% (n=5)

Backsliding 27.7% (n=36)

1->2	1->5	1->6	1->7	1->8	2->3	2->5	3->5	3->7	4->5
3.1% (n=4)	3.1% (n=4)	0.8% (n=1)	1.5% (n=2)	1.5% (n=2)	0.8% (n=1)	0.8% (n=1)	2.3% (n=3)	0.8% (n=1)	1.5% (n=2)
5->6	5->7	5->8	5->9	6->8	6->9	6->10	7->10	8->9	
0.8% (n=1)	3.1% (n=4)	2.3% (n=3)	0.8% (n=1)	0.8% (n=1)	0.8% (n=1)	0.8% (n=1)	1.5% (n=2)	0.8% (n=1)	

Table 1. Change in the coding and ranking of student responses (con't)

<table>
<tr><td></td><td>Code</td><td>Description</td><td>Rank</td><td>Pre Instruction</td><td colspan="2">Post Instruction</td></tr>
<tr><td colspan="7">Question 6b: Defend your answer with examples.</td></tr>
<tr><td rowspan="6">Change in Code
Pre->Post Instruction</td><td>1</td><td>Reflective use of example from MPU</td><td rowspan="2">Most</td><td>0.0% (n=0)</td><td colspan="2">8.5% (n=11)</td></tr>
<tr><td>2</td><td>Reflective use of other scientific example</td><td>33.8% (n=44)</td><td colspan="2">24.6% (n=32)</td></tr>
<tr><td>3</td><td>Vague reference to example from MPU</td><td rowspan="2"></td><td>0.0% (n=0)</td><td colspan="2">0.0% (n=0)</td></tr>
<tr><td>4</td><td>Vague reference to other scientific example</td><td>2.3% (n=3)</td><td colspan="2">2.3% (n=3)</td></tr>
<tr><td>5</td><td>Invalid reference/non-scientific example</td><td rowspan="2">Least</td><td>1.5% (n=2)</td><td colspan="2">0.8% (n=1)</td></tr>
<tr><td>6</td><td>Nonsense/no example</td><td>62.3% (n=81)</td><td colspan="2">63.8% (n=83)</td></tr>
<tr><td rowspan="9">Change in Rank
Pre->Post Instruction</td><td colspan="6">Improvement 18.5% (n=24)</td></tr>
<tr><td>5->1</td><td>6->1</td><td>4->2</td><td>5->2</td><td>6->2</td><td>6->4</td></tr>
<tr><td>0.8% (n=1)</td><td>6.2% (n=8)</td><td>0.8% (n=1)</td><td>0.8% (n=1)</td><td>9.2% (n=12)</td><td>0.8% (n=1)</td></tr>
<tr><td colspan="6">No Change 61.5% (n=80)</td></tr>
<tr><td>1->1</td><td>2->2</td><td>6->6</td><td>2->1</td><td>6->5</td><td></td></tr>
<tr><td>0.0% (n=0)</td><td>13.8% (n=18)</td><td>45.4% (n=59)</td><td>1.5% (n=2)</td><td>0.8% (n=1)</td><td></td></tr>
<tr><td colspan="6">Backsliding 20.0% (n=26)</td></tr>
<tr><td>2->4</td><td>2->6</td><td>4->6</td><td></td><td></td><td></td></tr>
<tr><td>1.5% (n=2)</td><td>16.9% (n=22)</td><td>1.5% (n=2)</td><td></td><td></td><td></td></tr>
</table>