ADDRESSING INSTRUCTIONAL CHALLENGES IDENTIFIED BY TEACHING ASSISTANTS TO IMPROVE OUTCOMES FOR UNDERGRADUATE STUDENTS IN AN INTRODUCTORY BIOLOGY LAB

BY

JESSICA B. MARTIN

A Thesis Submitted to the Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

Biology

May 2016

Winston-Salem, North Carolina

Approved By:

A. Daniel Johnson, Ph.D., Advisor

Brian W. Tague, Ph.D., Chair

Wayne E. Pratt, Ph.D.

ACKNOWLEDGMENTS

I would first like to thank my master's advisor, Dr. Dan Johnson for assisting me throughout the process of designing, implementing, analyzing and writing up this project. He spent countless hours meeting with me to discuss specific elements of this project and reviewed several drafts of this paper to offer advice about how to strengthen it. I would also like to thank Dr. Brian Tague and Dr. Wayne Pratt for the time they put into helping me at each stage of this project. Dr. Sarah Fick, my professor for a science pedagogy course, has also been a great mentor in pointing me to helpful resources and data analytics tools. Even after I had finished taking her course, she took the time to meet with me and offer advice about my project. I would also like to thank my husband Matthew, my twin sister Tiffany, my dad, my father-in-law Tim, and my friend Brittany for their willingness to carefully read and copy edit this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Introduction

Methods

Results

LIST OF ABBREVIATIONS

BASX	Biology Assessment Exam
BIO113	Ecology and Evolutionary Biology course
BIO114	Comparative Physiology course
HOTS	higher-order thinking skills
NOS	nature of science
NSSE	National Survey of Student Engagement
SB-NAS ver.1	student based needs assessment survey version 1
SB-NAS ver.2	student based needs assessment survey version 2
TA	graduate student teaching assistant
VNOS	Views of the Nature of Science assessment instrument
VOSTS	Views on Science-Technology-Society assessment instrument

ABSTRACT

This study explored learning outcomes in the biology department of a private, suburban, liberal arts institution with an undergraduate enrollment of about 4,800 students. Graduate student teaching assistants (TAs) are the primary instructors in four introductory biology labs. These TAs routinely report that undergraduate students in their labs are not meeting course objectives, but it has been unclear which objectives students miss and how frequently this occurs. It is also unknown whether these “gaps” in student learning are primarily due to (1) misconceptions or (2) holes in the knowledge and skills of undergraduate students or similar problems with TAs. This study aimed to address these issues by giving TAs a scenario-based survey on instructional challenges that asked them to rate how frequently they saw various gaps in student learning. This information was used to design an instructional plan that addressed these gaps and supported TAs by providing them with relevant content and pedagogical knowledge. This study showed that student understandings’ of the nature of science (NOS) increased, and students’ acceptance of responsibility for their work and performance decreased in the course in which the instructional plan was implemented. The impact on students’ higher-order thinking skills (HOTS) was inconclusive, likely due to differences in how students were prepared for HOTS assessments. Future improvements to the instructional plan should include adding more explicit instruction in three areas: key elements of the NOS, effective study methods, and the HOTS assessments used in the course.

INTRODUCTION

Similar to post-secondary faculty, graduate student teaching assistants (TAs) are faced with the challenge of teaching undergraduate students, especially in STEM fields (Sundberg, Armstrong, & Wischusen, 2005). At many institutions, TAs teach most laboratory sections and often teach some or all of lecture courses (Gardner & Jones, 2011). Graduate students often become TAs in their first year of graduate school, some right after earning their bachelor's degrees (Luft et al., 2004). New TAs often receive little or no pedagogical training. Just over half of TAs from a variety of disciplines reported having the opportunity to learn about teaching in their discipline specifically; the rest reported that no such training course was available (Golde & Dore, 2001).

Even when TAs are offered training, they do not always find it helpful. A study conducted at the University of Texas-Austin found that TAs felt their department's training program was not helpful because it did not cover information on laboratory instruction, grading lab reports or developing assessments (Luft et al., 2004). Another study reported that some students felt an hour long weekly training session on how to effectively teach science was too focused on strategies for teaching K-12 students versus strategies for teaching undergraduate students. When the session was expanded to two hours and more focus given to teaching undergraduates, students reported positive effects and the instructor reported a greater depth and application of teaching strategies in a course assignment (creating a lesson plan) near the end of the semester (Baumgartner, 2007).

Effective TA training is necessary because many TAs enter graduate school with misconceptions about and gaps in content and pedagogical knowledge. Barry and Dotger (2011) found that many TAs have content related questions that they do not ask in laboratory preparation sessions. Aquirre, Haggery, and Linder (1990) found that 33 of 74 pre-service teachers had naïve views of the NOS and that their views about how science worked influenced their beliefs about teaching. Many of the pre-service teachers involved in this study possessed undergraduate science degrees. It is, therefore, plausible that many incoming science graduate students, who have the same educational background, have similar misconceptions. For this reason, identifying and addressing graduate students' misconceptions should be a part of training them to teach undergraduate biology courses.

This study focused on the biology department of a private, suburban, liberal arts institution with an undergraduate enrollment of about 4,800 students, where graduate student teaching assistants are the primary instructors in four introductory biology labs. These TAs routinely report that undergraduate students in their labs are failing to meet course objectives, but it has been unclear which objectives students miss or how frequently this occurs. In addition, it is unknown whether these "gaps" in student learning are primarily due to misconceptions or gaps in the undergraduate students or their graduate student instructors. This study aimed to address these issues by giving TAs a scenario-based survey on instructional challenges that asked them to rate how frequently they encountered various gaps in their students' learning. This information was used to design an instructional plan that addressed these gaps and TA training materials to provide the TAs teaching the course with the relevant content and pedagogical knowledge needed to implement the instructional plan.

Defining the nature of science and its importance

The process by which scientific knowledge is developed, commonly referred to as the nature of science (NOS), is a crucial part of science that distinguishes it from other fields of study (Lederman & Zeidler, 1987). As described by Lederman et al. (2002), the NOS includes a variety of components. It is empirical, but all experiments are not composed of a fixed series of steps, and not all scientific theories can be directly tested. Instead, theories, which are important in driving scientific research, are supported by large amounts of indirect evidence. As this implies, scientific knowledge is also not entirely objective, as creativity, cultural and social factors influence theories. These factors are part of what makes scientific knowledge dynamic, another defining feature of the NOS.

Many college students have misconceptions about some foundational NOS concepts (Stein, Larrabee, & Barman, 2008), and these misconceptions affect their ability to learn science. For example, 57% of first-year college students believe that science has led to at least some definite right answers (Smith & Wenk, 2006). If students do not grasp that science does not lead to definite answers, then they will not see the value of repeating an experiment or will believe that any results that contradict current theories are invalid. This is the reason why the misconception that science leads to right answers that will never change, and other misconceptions about the NOS, need to be addressed. The National Research Council's report "How Students Learn" (2005) listed engaging student preconceptions and replacing misconceptions as critical to students grasping new information.

In this study, I assessed student understanding of the nature of science and probed for existing misconceptions using two previously published and validated instruments. These instruments are described below.

Views on Science-Technology-Society and Views of Nature of Science

The Views on Science-Technology-Society (VOSTS), as the name suggests, is a survey that can be used to determine students' perceptions of the relationships between science, technology, and society. It is composed of a variety of multiple-choice questions, with answer choices empirically-derived from Canadian high school students' responses to related open-ended questions. After the researchers composed multiple-choice options, subsets of the questions were given to groups of students in several rounds of testing. Depending on the round, these students were asked to verbally answer the questions or were interviewed after answering the questions to make sure the multiple-choice options matched students' views and to check each question's clarity and layout. Five thousand students were then given the revised questions and choice options; answer choices that were not selected often were eliminated. In total, the process of building, revising and validating the survey took six years (Aikenhead & Ryan, 1992).

The Views of Nature of Science (VNOS) questionnaire is a collection of open-ended questions originally developed by Lederman and O'Malley (1990). However, follow-up interviews revealed that researchers had not accurately inferred student views on the nature of science from answers to three of the seven questions. As a result, VNOS was revised and expanded, resulting in VNOS Form-B and VNOS Form-C (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). VNOS Form-C contained versions of five questions from VNOS-B and five new questions to more fully assess students' views of science and the scientific method (Norm G. Lederman et al., 2002). A team of science educators, scientists, and a science historian revised these questions and composed follow-up questions to further probe students' views on the nature of science in individual interviews (Abd-El-Khalick, 1998). In studies that administer VNOS Form-C, post-survey interviews serve to validate the survey by revealing whether students' answers to the 10 items match the views of science they express when asked follow-up questions.

Student responsibility and its importance

Student responsibility is a shorthand phrase that refers to a student mindset of accepting that they as learners are more responsible than teachers for certain learning tasks or outcomes (Zimmerman & Kitsantas, 2005). It is important for universities to help students accept responsibility for their schoolwork so that students will be motivated to put in the time and effort to effectively learn. Students who do not accept responsibility for their learning outcomes are more likely to think they lack control, and that their instructors are primarily responsible for learning success. When students perceive instructors to be more responsible for their learning, then they will not reflect on their mistakes and make personal efforts to improve. In fact, it has been found that the amount of effort students who do not feel responsible for their work put into assignments is dependent on how much their instructor holds them accountable (Lauermann & Karabenick, 2011). In this way, when students do not accept responsibility for their work, the work they put into learning becomes dependent on their instructor.

It is important that students put effort into studying and completing their class assignments. Obviously, effort correlates with higher academic performance. However, there is also an indirect link between students' study methods and perceived academic responsibility. Zimmerman and Kitsantas (2005) conducted a study of 179 girls at a highly selective academic school and found that those who engaged in high-quality study methods exhibited high self-efficacy for learning (that is a belief that they had the ability to regulate their learning). Examples of high-quality or effective study methods include forms of retrieval practice like self-quizzing or studying flashcards, or spaced out practice (Brown et al., 2014). Study methods like rereading text have been found to be less effective (Brown et al., 2014). Zimmerman and Kitsantas (2005) also found that students who exhibited a high self-efficacy for learning earned higher grades and were more likely to think learners were more responsible for certain learning tasks and outcomes than teachers (Zimmerman & Kitsantas, 2005). Fishman (2014) similarly found that students who felt responsible for their work were more likely to engage in deep learning.

Higher-order thinking and its importance

Higher-order thinking can be described as any type of thinking that involves more than recalling information. It can include many different types of skills including interpretation, analysis, evaluation, inference, and explanation (Facione, 1990). Many taxonomies have been created to differentiate between types of lower- and higher-order thinking. One of these taxonomies, Bloom's revised taxonomy (Anderson & Krathwohl, 2001), breaks up types of thinking by the type of knowledge and skills involved, referred to as the knowledge and cognitive process dimension respectively. According to this taxonomy, higher-order thinking is any type of thinking that requires more than the remember and understand categories of the cognitive process dimension (Anderson & Krathwohl, 2001).

Many college faculty, business leaders, and educational leaders agree that higher-order thinking is important. The Association of American Colleges and Universities released a report about the importance of universities teaching students critical thinking skills, among other learning skills (Association of American Colleges and Universities, 2005). This report also stated that at the time, 93% of university faculty saw two types of higher-order thinking (analytical and critical) as important goals of higher education. Employers agree: A 2013 survey by Hart Research Associates reported that 82% of employers want colleges to focus more on critical thinking and analytical reasoning skills; 93% care more about a job candidate having critical thinking, communication and problem-solving skills than his or her undergraduate major (Hart Research Associates, 2013).

Despite the value faculty and employers place on higher-order thinking, many undergraduate students are not developing them. The Commission on the Future of Higher Education reported in 2006 that many college graduates lacked expected thinking skills (Commission on the Future of Higher Education, 2006). Possible reasons why students are not gaining these skills even though faculty consider them important include students not considering higher-order thinking skills important, not recognizing when they are needed, and not keeping track of their thinking process (Halpern, 1998). In general biology courses, failure to develop higher order skills could also be due to a lack of explicit instruction. The 2010 National Survey of Student Engagement (NSSE) supports this idea. Seniors who completed the survey and majored in general biology reported that biology faculty spent almost 60% of class time lecturing and emphasized memorization in their classes (National Survey of Student Engagement, 2010).

Contributing to the problem of students not developing higher-order thinking skills in college is the lack of emphasis placed on it in K-12. The focus of many K-12 institutions is student performance on state standardized, high stakes tests (Smith & Szymanski, 2013). In addition, many colleges require students to take the SAT or ACT and use these scores in admissions decisions. U.S. News & World Report also requires SAT or ACT scores to rank colleges ("How U.S. News Calculated the 2015 Best Colleges Rankings - US News," n.d.), which likely discourages universities from dropping testing requirements.

The emphasis on standardized tests in K-12 schools and college admissions is problematic partly because all of these tests consist predominately of multiple-choice questions. A recent study found that even when students completed multiple-choice tests that included questions requiring higher-order thinking, they perceived higher-order thinking skills to be less important than students taking tests with multiple-choice and constructed response questions (Stanger-Hall, 2012). In addition, Stanger-Hall (2012) found that students taking tests with only multiple-choice questions displayed less cognitively active study behaviors and did not perform as well on a cumulative final exam. This suggests that the amount of standardized, multiple-choice tests students are required to take before college leads students to adopt passive study behaviors and view higher-order thinking skills as unimportant; a view they will likely carry with them when they enter college or the workforce.

For students to see the value of and develop higher-order thinking skills, teachers need to design instructional activities that explicitly require and encourage it. Well-developed instructional experiences have a demonstrated benefit. The 2013 NSSE report stated that undergraduate seniors in classes that emphasized higher-order thinking skills felt positively challenged twice as often as their peers in classes that did not stress such skills (National Survey of Student Engagement, 2013). But classes do not have to emphasize higher-order thinking skills to help students build these skills. The first cohort of a study that assessed students' skill development and collected background information on the individual and their experience at the university at the beginning and end of their college years found a connection between higher-order thinking skills and organized instruction. The study, which included approximately 2,200 students from 19 institutions, found that students who exhibited higher levels of reasoning and problem-solving skills reported experiencing high levels of clear and organized instruction. This was determined by a survey that asked about course elements like how class time was used, the clarity of explanations, use of examples, etc. (Pascarella & Blaich, 2013).

There are many different ways that instructors can help students develop higher-order thinking skills. Smith (1977) found that student thinking skills increased when students frequently interacted with their peers and participated in class activities at a high cognitive level. Likewise, Kim et al. (2013) found that group projects and scaffolding of activities and assignments increased thinking skills in an introductory science class. Another study found that writing and rewriting papers that required critical analysis, and class discussions that promoted students forming opinions, increased critical thinking (Tsui, 2002). Guided-inquiry based science labs have also been found to increase the development higher-order thinking skills compared to traditional lab instruction (Gupta, Burke, Mehta, & Greenbowe, 2015). As these studies show, instructors can increase students' higher-order thinking skills by incorporating one or more of a host of different activities. Instructors have flexibility in selecting activities that fit with the content they intend to cover and the structure of their class.

The purpose of this study

The purpose of this study was to identify whether constructing an instructional plan to address student understanding of the nature of science, acceptance of responsibility for their work and performance in the course, and development of higher order thinking skills (frequently identified problems identified by science teaching assistants at the institution where this study was conducted) would improve student performance in the targeted areas. It was predicted that the instructional plan would improve student performance in targeted areas more than traditional lab instruction, which was determined by comparing student performance in the lab in which the study was implemented to another introductory biology lab.

METHODS

Participants

Participants in this study were enrolled in an introductory level major's biology course entitled "Comparative Physiology" or BIO114. This is a survey course that introduces students to organismal biology and general physiology. Most students who take the class are first- or second-year students. Demographic information is shown in Table I. The course is one of four required classes that all biology majors must complete, and it fulfills part of an institutional divisional requirement of two math or natural science courses. The course consists of either three, 50-minute or two, 75-minute lectures and one, three-hour lab per week. Lecture sections are taught by multiple faculty members and range in size from 30-60 students. Lab sections of up to 16 undergraduates are taught by graduate student teaching assistants (TAs).

Gender	Class Year	Ethnicity
85 females	17, 1st-year	96, white
42 males	71, 2nd-year	12, black
	37, 3rd-year	15, Asian
	2, 4th-year	2 Hispanic/Latina(o)

TABLE I: Demographic information of all students in BIO114 in Fall 2015 who consented to participate in this study.

Laboratory coursework comprises 25-30% of students' overall course grade. Student grades are based on performance in six categories, listed below with typical percentage breakdown:

- Quizzes: 15%
- Homework assignments: 10%
- Participation: 10%
- Lab notebooks: 25%
- Oral presentation: 10%
- Lab reports: 30%

The lab course consists of six units: structure-function relationships, insect hormones, technical communication, nerve and muscle, metabolism, and transpiration. All units, except the technical communication unit, are two-stage inquiries: in the first week students learn how to use equipment and are introduced to the model system. In the second week, they design and conduct experiments. In the insect hormones unit, students present the results of their experiments as a group in an oral presentation. In the nerve and muscle and metabolism units, students write about the results of their experiment in individually written lab reports. Students also read and discuss four published scientific papers during the semester.

The comparison for this study was undergraduate students enrolled in an introductory level major's biology course entitled "Ecology and Evolutionary Biology" or BIO113. This course introduces students to basic principles of genetics, ecology and evolution. Most students who take the class are first- or second-year students. Demographic information is shown in Table II. Similar to "Comparative Physiology", this course is required for all biology majors, and it fulfills part of an institutional divisional requirement of two math or natural science courses. The course consists of either three, 50-minute or two, 75-minute lectures and one, three-hour lab per week. Lecture sections are team taught by two faculty members and range in size from 30-60 students. Lab sections of up to 18 undergraduates are taught by graduate student teaching assistants (TAs).

Gender	Class Year	Ethnicity
37 females	12, 1st-year	50, white
24 males	26, 2nd-year	5, black
	12, 3rd-year	4, Asian
	9, 4th-year	2 Hispanic/Latina(o)

TABLE II: Demographic information of students enrolled in BIO113 in Fall 2015 who consented to participate in this study.

Laboratory coursework typically comprises 25% of students' overall course grade. Student grades are based on performance in six categories, listed below with the percentage breakdown:

- Quizzes: 15%
- Homework assignments: 15%
- Participation: 10%
- Literature Discussions: 5%
- Lab summaries: 40% (15% for the first and 25% for the second)
- Oral presentation: 15%

The lab course consists of the following units: genetics review, ecological genetics, phenotypic variation, plant allocation and interaction, phylogeny and biodiversity, animal behavior, population modeling, and landscape ecology using satellite imagery. Three units are two-stage inquiries: in the first week students learn how to use equipment and are introduced to the model system. In the second week, they design and conduct experiments. Subsequently, students write a short summary of their experiment in the format of a scientific paper.

Statistical analyses

Unless otherwise noted, all statistical analyzes were conducted using R (R Development Core Team, 2008). Specific packages used are listed where relevant. Microsoft Excel was used for general data cleanup and visualization.

Institutional oversight and data security

The current study was submitted to WFU's Institutional Review Board for Human Subjects Research and approved on 8/13/2015 (Approval ID# IRB00022146).

Unless otherwise noted, surveys, quizzes, and other assessments were administered using WFU's Qualtrics Research Suite, via WFU site license. Qualtrics has been approved by WFU for data collection for human subject's research. To maintain data security, Qualtrics uses Transport Layer Security (TLS) encryption (also known as HTTPS) for all transmitted data. Surveys are protected with passwords and HTTP referrer checking. Third party data centers are SSAE-16

SOC II certified. All data at rest are encrypted, and data on deprecated hard drives are destroyed by U.S. DOD methods and delivered to a third-party data destruction service. Qualtrics meets the requirements of the FISMA Act of 2002 and meets or exceeds minimum requirements as outlined in FIPS Publication 200.

Data collected via Qualtrics Research Suite were downloaded as CSV files, and then either converted to Excel XML format or imported into R for further analysis.

Student based-needs assessment survey

In the first step of this project, science TAs were given a survey that asked them to identify where they most frequently saw students struggling in their courses. These areas are hereafter referred to as "gaps" because they are areas in which student performance and attitudes did not match instructor and course expectations. The TA survey hereafter called the scenario-based needs assessment survey (SB-NAS) was developed by the author and distributed to TAs in chemistry, physics and biology departments at a private, suburban, liberal arts institution with an undergraduate enrollment of about 4,800 students. The survey consisted of 16 scenario-based questions on eight specific gaps (two questions per gap) in higher-order thinking skills, conceptual content knowledge, or dispositions towards learning and self-regulation. These gaps were selected based on the experiences of the author and a survey of current areas of study within the field of biology education research. The eight student gaps assessed by the survey were: self-efficacy, motivation, organization of biology knowledge (ability to see the big picture), ability to transfer knowledge, acceptance of responsibility for their work and learning outcomes, development of higher-order thinking skills, understanding of the nature of science, and ability to work in groups.

The survey also included nine general information questions about the graduate student teaching assistants (TAs), including what science department the TA was a member of, how long they had been a TA, and whether they intended to pursue a teaching position in the future.

Version one of the survey (SB-NAS version 1) was administered before the Spring 2015 semester. Questions on this survey can be found in Appendix A.

TA responses to the survey were collected anonymously. Responses and feedback from respondents on SB-NAS ver.1 were used to revise the survey and create the second version, SB-NAS version two. Specifically, SB-NAS ver.2 was constructed to account for differences in the level of control and responsibility different TAs have over the lab courses in different departments (biology, chemistry, physics) and to make the questions discipline-neutral by removing subject specific content from the question stems. SB-NAS ver.2 can be found in Appendix B. Reliability of both surveys was assessed using McDonald's Omega in the MBESS package in R (Kelly & Lai, 2012).

Curriculum design and TA training

An initial instructional plan was designed in Fall 2014 to address three of the gaps TAs reported most regularly encountering in SB-NAS ver.1. These gaps were: student understanding of the nature of science, student development of higher-order thinking skills and student acceptance of responsibility for their learning. Consequently, the goals of the instructional plan were 1) to help students view the nature of science as fluid, question driven and evidence-based, 2) increase students' higher-order thinking skills, and 3) to help students accept responsibility for their learning (work and performance) in the course. One TA implemented the initial instructional plan in two sections of BIO114 lab in Spring 2015. Data related to each gap were collected, analyzed by the author and used to make the revised instructional plan. Six biology TAs, five in their first semester of teaching and the author of this study (who had implemented the initial instructional plan in Spring 2015), implemented the revised instructional plan in BIO114 lab in Fall 2015. The demographic information of these TAs can be found in Appendix C.

The instructional plans were developed using backwards design (Wiggins & McTighe, 1998). General learning objectives were established for the overall course and specific objectives for each unit or topic. Assessments and class activities were planned, each designed to address specific learning objectives as suggested by Anderson & Rogan (2011). Assessments included questions that assessed both content knowledge and skill development, which were both included in the course's general learning objectives. How the revised instructional plan aimed to improve and assess student performance related to each of the targeted student gaps is described in subsequent sections. After the revised instructional plan was created, training materials were developed to prepare new TAs to implement the instructional plan.

All TAs who were new to the university and participating in this study attended presentations and workshops that started 10 days before the Fall 2015 semester. The first set of training sessions and workshops introduced TAs to evidence-supported teaching practices including incorporating active learning into a curriculum and providing students with feedback. It also introduced students to general teaching principals like classroom management and the process of scientific teaching (Handelsman, Miller, & Pfund, 2006).

The second and final training session occurred a week later. It focused on the format of biology labs and detailed the revised instructional plan BIO114 TAs would be teaching. This interactive session began with a definition of inquiry-based learning and a discussion of what it meant to teach in this way and the benefits of doing so. This was followed by short lectures and interactive activities focused on the three gaps the BIO114 instructional plan was designed to address. These lectures and activities presented TAs with a description of each gap, how the curriculum was designed to address each gap, and a summary of how each student's performance or attitudes related to each gap would be assessed throughout the semester. After these initial training sessions, TAs attended weekly meetings in which they were given information on how to conduct laboratory experiments, work with students and grade student assignments.

Information on specific gaps

Data from the initial instructional plan were only used to plan the revised instructional plan and are not included in this thesis. When the revised instructional plan was administered in Fall 2015, students in BIO113 or BIO114 were asked to consent to participate in this study. Data on consented students' achievement and attitudes related to the three specific gaps addressed by the instructional plan were collected throughout the semester but not analyzed until after final grades had been submitted. For data analysis, a second investigator generated 300 pseudonyms using 300 non-repeating first names (150 each of male and female) and 300 last names, selected by a random name generator for fiction authors (https://www.randomlists.com/) to represent a range of ethnicities. The second investigator then assigned each consenting student participant a random 8-digit ID number and pseudonym. Original participant names were scrubbed from all documents, survey responses, etc., and replaced with corresponding assigned pseudonyms. The lead study author had no access to the identity key list at any time during analysis, only to pseudonyms or randomized ID numbers. When individual responses are mentioned in this paper, they are referred to by their assigned pseudonym last name and the course in which the student was enrolled.

Nature of science (NOS)

Explicit instruction on the NOS began the third class session when students completed a guided activity (this was the second instructional class because students completed an internal student assessment exam during the first class session). As a part of this activity, students visited a website and answered questions about the process of science and scientific testing, with a special emphasis on the purpose of experiments ("Understanding Science," 2016). Students completed this activity in groups of three to four, after which the TA led a class discussion about what they found. TAs were instructed to emphasize the following points during this discussion: 1) the process of science is not predetermined or linear, 2) scientific knowledge is empirically derived, and 3) in an experiment it is important to control all variables other than the one being tested. This guided activity and discussion were incorporated into the curriculum because adding explicit instruction to inquiry-based activities has been found to increase student understandings of the NOS (Khishfe & Abd-El-Khalick, 2002).

In addition to the activities described above, the nature of science was modeled to students throughout the semester by their TAs. TAs led discussions of four scientific literature articles in which they guided students in learning how to critique an experiment and assess the reliability and shortcomings of a research study. Article discussions were included in the instructional plans because it has been previously reported that doing so can advance student understanding of the NOS (Hugerat & Kortam, 2014; Wenk & Tronsky, 2011). Foundational aspects of the NOS, particularly that science is fluid, question driven and evidence-based, were reiterated throughout the semester as students completed the inquiry-based labs that were already built into the curriculum.

Student understanding of the NOS was assessed at the beginning and end of the semester using selected questions from two previously published and validated instruments, Views on Science-Technology-Society (VOSTS), and the Views of the Nature of Science Questionnaire (VNOS), forms B and C (Aikenhead & Ryan, 1992; Lederman et al., 2002). The 10 VOSTS and two VNOS questions used in this study can be found in Appendix D. These questions were administered twice, first as a part of the biology department's internal student assessment exam (Biology Assessment Exam or BASX; pre-test) and again at the end of the semester in a final lab evaluation (post-test). All students enrolled in BIO113 and BIO114 completed the BASX exam on the first day and final lab evaluation on the last day of lab.

VOSTS questions were analyzed in two ways. First, the number of BIO113 and BIO114 students who selected each answer choice on the pre- and post-test was determined. All answers were categorized as appropriate, plausible or naïve/null and the proportion of each type of answer choice was calculated for each question on the pre- and post-test. Second, the answer choices were converted to numbers with the categories appropriate, plausible, or naïve/null scored as 3 1/2, 1, or 0 respectively. This method of analysis was based on one of the arguments of Vazquez-Alonso & Manassero-Mas (1999) who argued that this method better reflects student understandings of NOS than other scoring methods. Wilcoxon signed-rank tests were used to compare pre- and post-treatment scores for BIO113 and BIO114. Mann Whitney U-tests were used to compare BIO113 and BIO114 scores pre- and post-treatment. Both Wilcoxon signed-rank tests and Mann Whitney U-tests were calculated in the coin package in R (Hothorn et al., 2015). The effect size of the BIO113 or BIO114 curriculum on students' beliefs about the nature of science, as evidenced by their answers to the VOSTS questions, was also determined by dividing the Z score by the square root of the sum of the number of samples in each group, as has been previous recommended (Pallant, 2007).

BIO113 and BIO114 student answers to the 2 VNOS questions were put in coding categories according to the terms they contained, ideas they described and the language used. Specifically, answers to the first VNOS question "What is an experiment?" were put in coding categories based on whether the answer mentioned the term "hypothesis" (or a similar term). Answers that did not mention hypothesis or a similar term were put in the coding category "not coded." Answers that mentioned the term hypothesis (or a similar term) were put in a coding category based on the language used to discuss hypothesis testing. Answers that mentioned hypothesis testing in absolute language (e.g. to determine if the hypothesis was correct) were put in the "absolute" category. Answers that mentioned hypothesis testing in non-absolute language (e.g. to determine if the hypothesis was supported) were put in the "non-absolute" category. Answers that used vague language that could not be classified as absolute or non-absolute were put in the "vague" category.

Student answers to the second VNOS question, "Does the development of scientific knowledge require experiments?" were put in coding categories based on whether the answer said experiments were or were not required. Answers that said experiments were not required were put in the "experiments not required" category. Answers that said experiments were required to obtain absolute knowledge (e.g. experiments are required to obtain facts) were put in the "absolute" category. Answers that said experiments were required to gather non-absolute knowledge (e.g. to obtain evidence for theories) were put in the "non-absolute" category. Answers that used vague language that could not be classified as absolute or non-absolute were put in the "vague" category. All the coding categories for VNOS questions one and two were inductively developed (created based on student responses) by the author from the data collected in Spring 2015 and revised based on the answers of Fall 2015 study participants. Examples of student answers that fell in each of the coding categories described above for VNOS questions one and two are shown in Table III.

What is an experiment?			Are experiments required?			
Vaguely mentions hypothesis testing	Mentions hypothesis testing in non-absolute language	Mentions hypothesis testing in absolute language	States experiments are not required	States experiments are required in absolute language	States experiments are required in non-absolute language	Vaguely states experiments are required
Sirois (BIO114) on pre-test: "An experiment is when you are testing a hypothesis using different controls and variables"	Gourd (BIO114) on post-test: "An experiment is testing a hypothesis and collecting data to see if you are able to support your hypothesis with said data."	Afanador (BIO113) on pre-test: "A test to know if a hypothesis is correct or not."	Graff (BIO113) on post-test: "Yes and no. Yes because it allows one to apply scientific knowledge. No because one can easily gain scientific knowledge through reading the textbook or lab manual."	Balis (BIO113) on pre-test: "Yes, because all scientific statements must be proven. Since experiments are designed to test hypothesis, if that hypothesis is tested through an experiment, then it can be considered scientific knowledge."	Falcone (BIO114) on pre-test: "Yes. Experiments are the only way to accurately determine scientific knowledge. If an experiment can be run over and over again and yield the same results, it is more likely to be true."	Halloway (BIO114) on post-test: "Yes, if you do not test the outcome of what is going on then you can not create theories and other scientific knowledge that other people go by."

TABLE III: Examples of student answers that fell in each of the main categories used to code VNOS questions one and two. All questions were placed in one of the categories.

Some BIO113 and BIO114 students mentioned scientific ideas or used scientific terms not directly related to the VNOS questions. Such answers were put in a second set of coding categories that grouped answers containing related terms or ideas. These coding categories were important to include because they provided evidence of past instruction in the pre-test and evidence of what was learned in the post-test and provided additional insight into students' views of the NOS. These coding categories were all created based on the answers of Fall 2015 study participants. After the codes had been created, all student answers were re-read, and those that mentioned a scientific idea or term other than those considered in the first set of coding categories were coded a second time. The coding categories used in this analysis were: gave a description of "science methods or procedures" (may not use the term scientific method), used the term "scientific method", mentioned the need for "experiments to be repeated", used the scientific term "variables", used the terms "theories or assumptions", and mentioned the idea that "scientific knowledge builds and/or changes". Examples of student answers that fell in each of the second set of coding categories is shown in Table IV.

Mentions scientific term or idea	Scientific methods (may not include term)	Scientific Method (term)	Variables (term)	Theories and assumptions	Need for repetition	Builds and changes
Thatch (BIO113) on post-test in response to question 1: "A test of a given concept through multiple trials of control vs. experimental groups."	Shock (BIO113) on pre-test in response to question 1: "An experiment is a process with specifically defined steps to test a hypothesis."	Micha (BIO113) on pre-test in response to question 1: "An experiment is an investigative series of tests used as a staple of the scientific method."	Duty (BIO114) on pre-test in response to question 1: "A procedure with a set of independent, dependent, and controlled variables established to confirm or disprove a hypothesis."	Hickel (BIO114) on pre-test in response to question 2: "Yes because without experiments, we would not be able to either support or disprove theories that have been created. "	Vanness (BIO114) on post-test in response to question 1: "a scientific study of a question (hypothesis). It involves testing and retesting."	Thiel (BIO113) on post-test in response to question 2: "Of course it does. Without experiments much of the knowledge that now know could have no basis. However scientific ideas have to also be liable to change."

TABLE IV: Examples of student answers that fell in categories related to different elements of the nature of science. Only answers with terms or ideas related to a category were coded. Some answers were placed into multiple categories.

Student responsibility

Instruction aimed at helping students accept responsibility for their work and performance in the course began the first week of class with an in-class assignment on effective study strategies. Students completed this assignment as individuals. TAs then led a discussion on the activity and were instructed to emphasize effective study strategies like retrieval and spaced practice during this exercise. The topic of this activity was selected based on the previously reported indirect link between study strategies and perceived responsibility for learning (Zimmerman & Kitsantas, 2005). TAs were instructed to remind students of these effective study strategies throughout the semester when students initiated discussions about how to improve their performance in the course.

Students' acceptance of responsibility for certain learning tasks was assessed using 11 multiple-choice questions, which can be found in Appendix D. Two of these questions were on tasks for which teachers are commonly accepted to be primarily responsible, while 9 were on tasks for which teachers would say students are primarily responsible. Like questions on the nature of science, these questions were included in the BASX (pre-test) completed on the first day of lab and in the end of course evaluation (post-test) completed on the last day of lab. Each question gave students a specific learning task (e.g. taking notes in class) and asked them to rate whether teachers or students were mainly, definitely or slightly more responsible or whether both were equally responsible for the given task. The type of questions and answer choices were based off an assessment created by Zimmerman and Kitsantas (2005). Wilcoxon signed-rank tests were used to compare pre- and post-test scores for BIO113 and BIO114. Mann Whitney U-tests were used to compare BIO113 and BIO114 scores pre- and post-treatment, and student responses to the 2 tasks teachers were predominately responsible for to the 9 tasks students were predominately responsible for. Both Wilcoxon signed-rank tests and Mann Whitney U-tests were calculated in the coin package in R (Hothorn et al., 2015.) The effect sizes of the BIO114 and BIO113 curriculum on student's beliefs about their responsibility for certain learning tasks and outcomes, as evidenced by their answers to the questions described above, were determined by dividing the Z score by the square root of the sum of the number of samples in each group.

In addition to the questions described above, BIO114 students' acceptance of responsibility for their work and performance was assessed throughout the semester via diagnostic learning logs. Each student was instructed to complete eight learning logs throughout the semester, two each on class sessions, quizzes, lab notebooks and lab reports. The instructions students were given on writing learning logs can be found in Appendix F. Students were instructed to complete one learning log for each type of assignment in the first half of the semester and a second for each type of assignment in the final half. These learning logs asked students to list what they understood and did not understand, identify errors and good answers they made on class assignments, and describe how they prepared for the assignment and planned to improve on future assignments. Learning log instructions, including the questions, were modeled after Angelo and Cross's (1993) description of diagnostic learning logs.

All learning logs completed by a random subset of eight students were used to develop coding categories. The descriptive codes created were: preparation, improve/learn more, mistakes, and grades. A description of each code and how the answers were separated within each code can be found in Table V.

After the coding categories had been developed, all learning logs completed by a random subset of 20 students taught by five of the six TAs involved in the study were analyzed. The learning logs of students taught by one TA were excluded because their students frequently completed learning logs on assignments other than those specified in the instructions. Due to significant changes in the way answers that discussed mistakes were sub-divided within the mistakes category, all answers or parts of answers that fell in this coding category were resorted into sub-categories in a third round of coding.

Descriptive Categories			
Preparation: This includes all comments made about how the student prepared for an assignment. These are divided up based on 1-whether or not they were all done by the student or with other students and/ or the instructor and 2-by whether or not the actions listed are specific or vague.	**Improve/Learn more:** This includes all actions the student says they will do in order to learn more about a subject or improve on future assignments. It is divided up the same way as preparation.	**Mistakes**: This includes all answers in which the student mentions not studying, answering a question well, making any other kind of mistake. These are divided up based on 1- whether the mistakes are explicitly stated, implied or blamed on other things and 2- by the type of mistake or how a mistake was implied.	**Grades**: This includes all answers in which the student implies or explicitly mentions losing points, not doing well, how to increase their grade, and/or earning a grade. They are divided up based on whether the answer implies an external or internal locus of control.

TABLE V: A description of what type of answers fit in each descriptive code and how answers were separated within each bin to account for whether or not the student was showing signs of accepting or not accepting responsibility for their work and performance on a given assignment.

Within coding categories, answers were divided by whether or not they suggested the student was or was not accepting responsibility. As a part of this process, the clearness of the connection between the student's answer and whether or not they were accepting responsibility was considered. In the "preparation" category, whether or not students were accepting responsibility was determined by whether they depended on someone else to prepare for assignments. Answers were then sub-divided by how clearly the answer indicted whether the student was relying on personal actions or the actions of others. For example, a student who listed specific actions for how they studied for a quiz likely perceived studying for a quiz to be their responsibility. However, a student that vaguely described how they prepared, like the student in Table VI who writes they "made sure they understood the material" may not be perceiving studying to be their responsibility. This student could, for example, perceive teachers as being responsible for making sure students understand the material. The categories and sub-categories used to classify if students who described how they prepared for an assignment were or were not accepting responsibility are shown in Table VI. This table also includes sample answers.

Self-driven preparation			Preparation based only on others	Mixed	
Took specific actions	Not clear how they prepared from action listed (vague actions given)	Mix of specific and vague actions	Only mentions talking/working with other students	Actions including talking/ working with other students	Mention actions incl. talking/ working with others (teacher and students or cannot determine)
Bade on 2nd lab report log: "...prior to writing the paper I read through many articles written about the same general topic to understand the biological processes behind explaining why metabolic rate was affected the way that it was by a lowered pH level."	Araiza on 1st quiz learning log: "...I made sure that I fully understood the material before the lab was completed..."	Cram on 1st lab report learning log: "...I accomplished this [writing the discussion well] by following the guidelines as well as making meaningful extra connections to understand the results."	Graff on 2nd lab notebook learning log: "...I met with my group and we sat down together to formulate an ideal experiment and how to express it in understandable words for the reader."	Glick on 2nd lab report learning log: "...Some things I did to strengthen my paper was to re-read it and have others peer edit it."	Graff on 2nd class session learning log: "...I got a better understanding [of the oxygen probe] by preparing for quiz 9. I also asked questions about the probe…"

TABLE VI: Examples of student answers that fell into categories related to how students prepared for assignments. Only answers or parts of answers with ideas related to preparation were coded.

Answers in the preparation category were also re-coded in a third round of coding. All the answers that mentioned study methods were categorized by whether or not the study methods were "effective" or "ineffective." The only effective study method mentioned in the learning logs analyzed described spaced out retrieval practice in which the student tested him/herself on the material s/he needed to know. The most commonly described ineffective study methods were rereading the lab manual or other course readings and rewriting information.

The "improvement" category was sub-divided in a way similar to the preparation category. Answers were first split by the type of actions the student described they would take to improve.

Students who described actions they would personally take to improve were more likely to be accepting responsibility for their performance on specific assignments and in the course as a whole. However, like with the preparation category, the strength of this association was also considered so vague actions that could be based on the instructor were separated from answers that described specific actions the student would take. These sub-categories, along with example answers, are shown in Table VII.

Self-driven improvement/learning			Improvement/learning based on others			Mixed		
Mentions specific actions	Mentions vague actions	Mentions mix of specific and vague actions	Only mentions talking to instructor	Only mentions talking/ working with other students	Only mentions talking/ working with others (not specified or someone other than instructor or students)	Includes talking to instruc-tor	Includes talking/ working with other students	Includes talking/ working w/ others (not specified or some-one other than ins-tructor, students)
Bade on 2nd lab notebook log: "...I could just make sure to do more outside research to prepare for the writing of the experi-ment…"	Baney on 2nd lab notebook log: "I could make sure I com-pletely under-stand the procedure and how it relates to other scientific work."	Eddinger 2nd lab notebook log: "Next time, I should try to find more primary literature and learn how to incor-porate it."	Earle on 2nd class session log: "I could go to my TA to ask for clarifi-cation, he studies trans-piration! I could also ask my lecture professor; he's a plant physio-logist!"	Gurr on 1st lab notebook log: "...I might have a member or my lab group peer review my notebook to check if any parts are miss-ing or any parts need improve-ment…"	Glick on 2nd lab report log: "...I would go to my TA and have my TA read over it before I hand it...I would also have another student in my lab group re-read my lab [report].... it might be interesting to go to the writing center and see what they have to say."	Castenada on 2nd class session learning log: "To get a better under-standing, I can read the Bio114 text book [sic]. I can also look it up on Youtube or online or talk to my TA or pro-fessor."	Bay on 1st lab report log: "...I think it would be effective to find more primary literature sources...I also think having more people to peer-review the paper..."	Carmon on 1st class session log: "The best way to gain a better under-standing... is to simply read multiple high quality reports... Another important thing to do is ask questions when I am strug-gling..."

TABLE VII: Examples of student answers that fell into categories related to how students planned to improve on future assignments. Only answers or parts of answers with ideas related to preparation were coded.

The “mistakes” category was sub-divided by whether or not the mistake was explicitly stated or implied and by the type of mistake students described making. Students who explicitly stated making a mistake were more likely to be accepting responsibility for making a mistake. These answers were sub-divided by the type of mistake the student admitted to making because this provided information about what types of tasks the student considers him/herself to be responsible for. Students who implied making mistakes may not have perceived their mistake as their responsibility and were thus separated from student answers that explicitly stated making a mistake. In addition, students who seemed unsure if they made a mistake or who gave reasons why they made a mistake were separated because these students were less likely to be accepting responsibility for their mistake. These sub-categories, along with example answers, are shown in Table VIII.

The answers that mentioned grades were placed in the “grades” category. Answers in this category were subdivided by whether or not they implied an external or internal locus of control, or whether this could not be determined from the answer. Students with an internal locus of control were more likely to be accepting responsibility for their work and performance in the course. These sub-categories, along with example answers, are shown in Table IX.

Explicitly admits they made mistakes			Implies they may have made mistakes			Other	
Did not do an assignment or at least one part of an assignment well OR answered a question incorrectly	Did not prepare/ study enough or made a mistake when doing assignment	Makes at least one connection of specific things done incorrectly or not done well because of a mistake or lack of knowledge	States things could/ should have been done better	States next time could improve by taking very specific actions. May state reasons why.	Lists areas needing improve- ment, reason did well, or how to do better next time in way that may imply a mistake was made	Seems unsure if they made a mistake	Gives excuses/ reasons why the parts of the paper or assignment were not done well
Declue on 1st quiz learning log: "I got this question wrong because my answer was not correct….I did not explain my reasoning well and... I never answered the question... "	Glick on 2nd quiz learning log: "I did not preform [sic] as well on other quizzes because I did not take notes on what my TA was saying..."	Castenada on 1st lab report learning log: "...I think I repeated myself too much because I wasn't really sure where I was going with what I was saying."	Earle on 1st lab notebook learning log: "...I could be more concise in my methods section, not necessarily outlining all of the details of the experi- ment...I also need to draw an explicit con- clusion about my hypothesis from my data..."	Batie on 2nd lab notebook: "...I will be more thorough in my reasoning with certain things like explaining the statis- tical signifi- cance of results... rather than just stating a p-value and not inter- preting it in context..."	Culton on 1st lab report: "I need to be more specific about why I made the predictions I made... I repeated some of what I said in the results in my discussion section, so I need to cut that out next time. In general I think my discussion was a little too long and I contra- dicted myself a little..."	Bay on 1st lab report learning log: "... while I thought each section was concise, I may have been able to shorten my methods section more...."	Batie on 1st lab report: "I probably could have been briefer in methods section. It was hard to get enough detail to make it replicable yet still have the section not seem lengthy. I would have liked to incorporate more sources into discus- sion, but found it difficult to synthesize ideas..."

TABLE VIII: Examples of student answers that fell into categories related to making mistakes. Only answers or parts of answers with ideas related to preparation were coded.

Mentions grade in a way that may imply external locus of control	Mentions grade in a way that may imply internal locus of control	Mixed (internal and external OR could be either)	
		Either/cannot tell	Internal and external
Glick on 1st lab notebook learning log: "...I did not realize that I needed to be formal in terms of my writing and that this lab notebook was considered a large piece of work...In addition, I needed to say instead of disproving our hypothesis I needed to say our hypothesis was not supported."	Enos on 2nd quiz learning log: "...On some quizzes that I haven't received an A, I often did not study the material beforehand, or simply rushed through the quiz."	Fitzsimons on 2nd quiz learning log: "... On previous quizzes I have not focused on the specifics in topics and focused more on the big picture ideas which did not help me as much on the quizzes. Because I focused more on the minute details of the lab report requirements I was better able to understand how to answer the questions being asked."	Araiza on 1st quiz learning log: "I don't think I performed as well on the first quiz because I did not know what to expect or how to prepare for it. That experience taught me what I should do for the next quiz."

TABLE IX: Examples of student answers that fell into categories related to grades. Only answers or parts of answers with ideas related to preparation were coded.

Higher-order thinking skills

Students were supported in developing higher-order thinking skills (HOTS) in reading and discussing scientific articles, which Hugerat and Kortam (2014) found increased the development of HOTS. Students also kept lab notebooks, which included developing explanations of their results and supporting their explanations like they would do later in the semester in more formal lab reports. In addition, students also completed short, non-graded activities that contained questions requiring HOTS at the beginning of every class. TAs were instructed to ask students questions that required HOTS while students completed these activities and to model HOTS during class discussions of these activities.

A baseline for HOTS was established early in the semester for BIO114 students using questions on quizzes students completed the first or second week of class. Students were then given the opportunity to practice and demonstrate HOTS on subsequent quizzes given almost weekly. A subset of quiz questions specifically requiring HOTS were used by all TAs, so all students were asked the same questions. From the middle to end of the semester, students were also asked to use HOTS to develop and support a physiological explanation of their results. The quiz questions assigned as a part of the revised instructional plan and lab report instructions regarding writing a physiological explanation can be found in Appendix G.

Students HOTS were assessed using a previously described method (McNeill & Krajcik, 2011) of breaking their answers down into claim, evidence and reasoning and scoring each element from 0-3. All three questions from quiz two were scored as an indicator of students' higher-order thinking skills at the start of the class. The second question from quiz six was scored as an intermediate measure and the physiological explanations in lab reports one and two as a second intermediate and final measure of students' higher-order thinking skills. These assignments were selected for analysis due to similarities in question type: all quiz questions and lab reports asked students to develop an explanation (make a claim) and support this claim with evidence and reasoning. The original tool described by McNeill and Krajcik's (2011) and tools constructed to score each assignment can be found in Appendix H. Students who did not complete all of the questions were excluded from the analysis. Scores for all questions were compared to each other using Wilcoxon signed-rank tests in the coin package in R (Hothorn et al., 2015).

RESULTS

Student based-needs assessment survey

SB-NAS ver.1 (Appendix A) was completed by 24 science TAs: 6 in the chemistry department, 17 in biology and 1 in physics. Coefficient omega was 0.77 with a 95% confidence interval of 0.43- 0.89. This means that the survey's internal consistency was in the range of unacceptable to good. The two questions on each instructional challenges were combined to determine how frequently each gap was reported. The gaps with the highest reported frequency were 1) student understanding of the nature of science, 2) student development of critical thinking skills, 3) student acceptance of personal responsibility, and 4) student organization of biology knowledge. The fourth gap was not selected to be addressed in the instructional plans because of concerns about the questions, which TAs could have interpreted to be about higher-order thinking skills. For this reason, higher-order thinking skills and not students' organization of biology knowledge, was selected to be addressed in the instructional plans.

SB-NAS ver.2 (Appendix B) was completed by 26 science teaching assistants: 10 in the chemistry department at Wake Forest University, 13 in biology and 3 in physics. Two of the 26 teaching assistants who completed the survey only answered half of the survey questions. These answers were included in the analysis because the TAs answered all questions for four of the instructional challenges. Coefficient omega was 0.90 with a 95% confidence interval of 0.71- 0.96. This means that the survey's internal consistency was acceptable, good or excellent. SB-NAS ver.2 contained slightly different versions of each question for each science department. The answers to the different versions of each question were combined for analysis. The two questions on each instructional challenge were also combined to determine how frequently each gap was reported. Gaps were reported as rare if the most commonly selected answer choice was never or not applicable and occasional if the commonly selected answer was 1 or 2 times a semester. Gaps were reported as frequent if the most commonly selected answer was three or more times a semester; gaps were reported as routine if the most commonly selected answer was almost weekly. As shown in Figure 1, gaps reported as occasionally encountered in SB-NAS ver.2 were 1) student self-efficacy, 2) student motivation and 3) student acceptance of personal responsibility. Gaps reported as frequently encountered were 1) student understanding of the NOS and 2) student ability to work as a group. Gaps reported as routinely encountered were 1) student organization of biology knowledge (seeing the big picture) and 2) student development of higher order thinking skills. Figures showing how frequently TAs in different departments reported encountering each instructional challenge is shown in Appendix E.

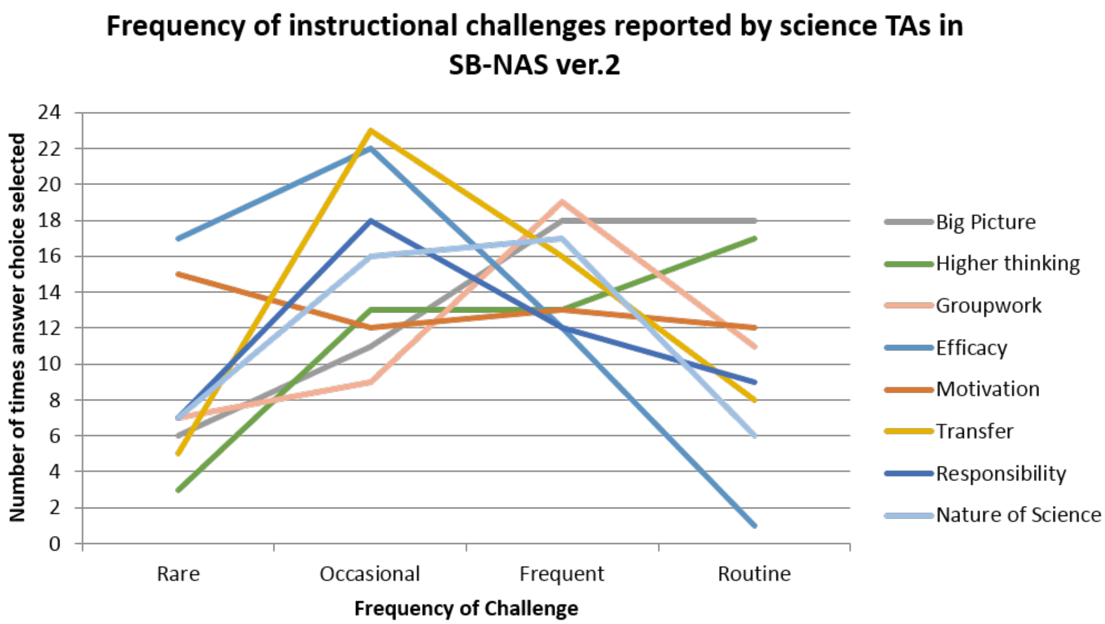

FIGURE 1: The frequency of instructional challenges reported by science TAs in SB-NAS ver.2 administered at the beginning of the Fall 2015 semester. The x-axis shows the frequency of each instructional challenge with rare representing the answer choices not applicable or never, occasional one or two times a semester, frequent three or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=46-52 times the answer choices was selected).

Nature of Science

Students in BIO113 and BIO114 did not answer the VOSTS questions significantly differently on the pre-test (Z=-0.7854 and p-value=0.4322) or post-test (Z=0.2531 and p-value=0.8002). Within each course, students also did not answer the VOSTS questions significantly differently and effect sizes were small to medium (BIO113: Z=0.4371 and p-value=0.662, and effect size=0.01600; BIO114: Z=1.9183 and p-value=0.05507, and effect size=0.4599).

There was no clear trend across questions of students changing from one type of answer (appropriate, plausible, naïve) to another type as shown in Figure 2. For example, notice in Figure 2 that appropriate answers were selected more frequently on the post-test compared to the pre-test for four questions (10112, 90521, 90611, and 90711). In contrast, appropriate answers were selected less frequently on the post-test compared to the pre-test for five questions (90411, 90621, 90631, 90811 and 91012). There was little change in the number of times appropriate answers were selected between the pre- and post-tests for one question (90511). This shows that students were not consistently picking a more appropriate or less appropriate answer choice across questions. A similar trend can be seen in the other answer categories. For more detail on how selected answers changed within each answer type, refer to figures in Appendix I that show the number of times each answer choice was selected by BIO113 students on the pre- and post-tests.

Only subtle changes in how BIO114 students selected each type of answer between the pre- and post-tests were observed. The proportion of each type of answer did not change for most questions between the pre- and post-tests as shown in Figure 3. The most common change in selected answers was from one plausible answer choice to another. Figures showing the number of times each answer choice was selected by BIO114 students on the pre- and post-tests can be found in Appendix I.

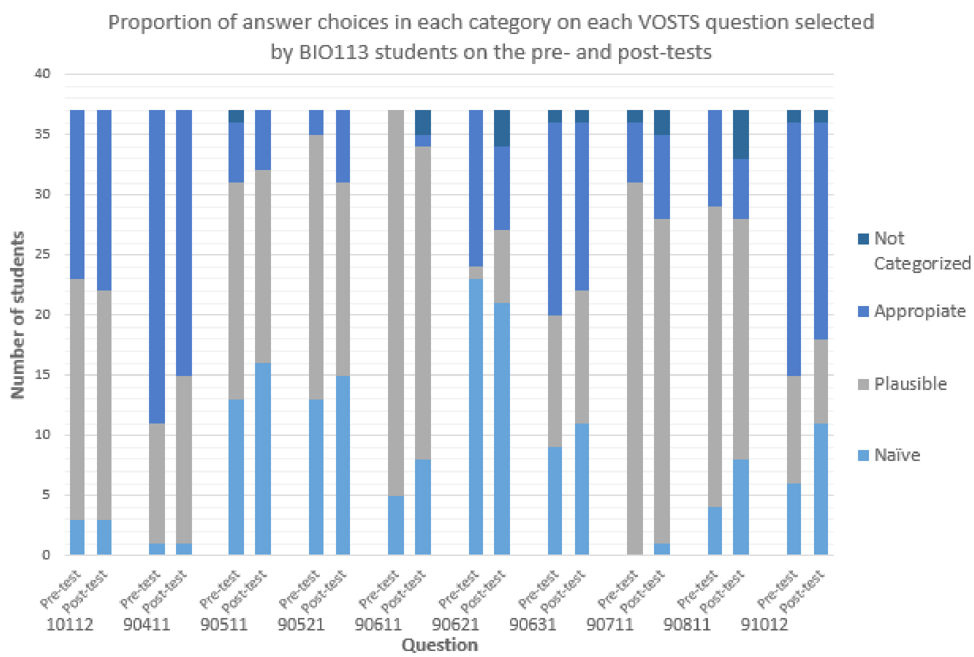

FIGURE 2: The figure above shows the proportion of each answer choice selected by BIO113 students (n=37) on the pre- and post-tests. Notice that changes in the proportion of each type of answer varies by question.

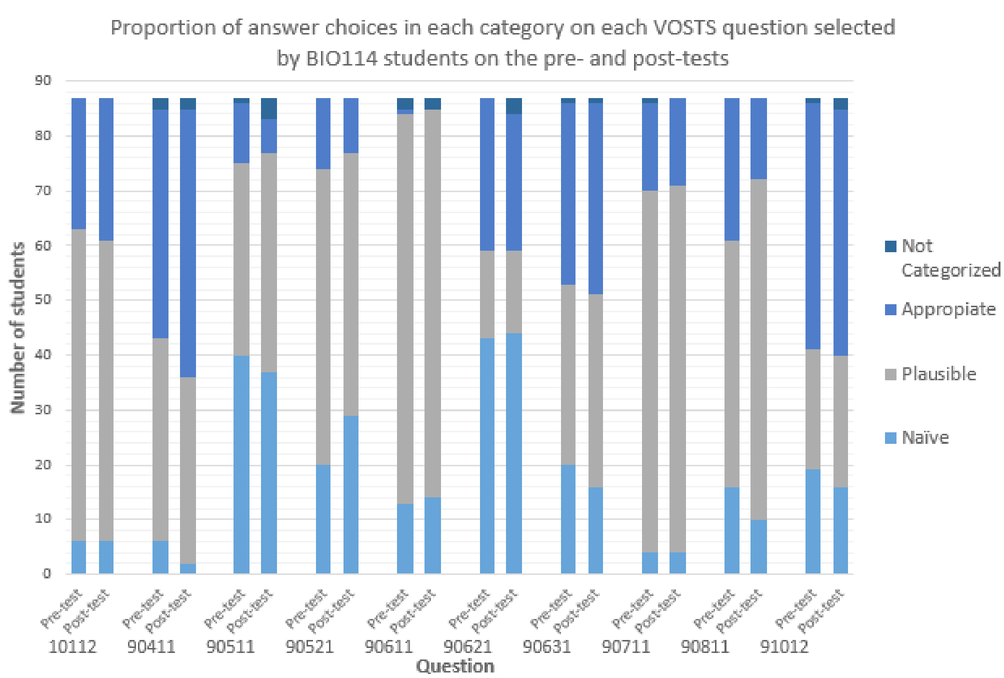

FIGURE 3: The figure above shows the proportion of each type of answer choice selected by BIO114 students (n=87) on the pre- and post-tests. Notice that the proportion of each type of answer is approximately the same between the pre- and post-tests for almost all questions.

Of the students who consented to participate in the study, 53 students in BIO113 and 111 students in BIO114 answered both questions from the Views on the Nature of Science survey on pre- and post-tests. The number of BIO113 and BIO114 student answers that fell in each coding category for VNOS question one, "What is an experiment?" is shown in Table X. All BIO113 student answers were coded but five BIO114 student answers on the pre-test were not coded because they did not mention hypothesis or a similar term. BIO113 students gave more vague answers and fewer answers containing absolute or non-absolute language on the post-test compared to the pre-test. BIO114 students also gave more vague answers and fewer answers containing absolute language but had more answers containing non-absolute language on the post-test compared to the pre-test.

		Vague language	Non-absolute language	Absolute language	Not coded
BIO113	pre-test	33	8	12	0
	post-test	44	4	5	0
BIO114	pre-test	63	8	35	5
	post-test	81	10	20	0

TABLE X: A table showing how many BIO113 and BIO114 student answers for VNOS question one were placed in each coding category on the pre- and post-tests. Note that on the BIO114 pre-test, five student answers were not put in any of the coding categories because they did not mention hypothesis or a similar term.

The number of BIO113 and BIO114 student answers that fell in each coding category for VNOS question two, "Does the development of scientific knowledge require experiments? If yes or no, explain why. Give an example to defend your position." is shown in Table XI. All BIO113 and BIO114 answers fit into a coding category. A similar number of BIO113 student answers fell in each category on the pre- and post-tests. BIO114 student answers on the pre-test were similar to BIO113 student answers. However, on the post-test BIO114 students said that experiments are required to obtain absolute knowledge about half as often as on the pre-test, with 25 student answers in this category compared to the 51 on the pre-test. The number of student answers in all other categories increased.

		Vague language	Absolute language	Non-absolute language	Experiments not required
BIO113	pre-test	16	18	15	4
	post-test	16	16	16	7
BIO114	pre-test	38	51	10	10
	post-test	46	25	23	17

TABLE XI: A table showing how many BIO113 and BIO114 student answers for VNOS question two were placed in each coding category on the pre- and post-tests.

In addition to coding the answers as described above, answers were also categorized if they mentioned other specific elements of the nature of science (NOS). The number of BIO113 and BIO114 student answers that included a specific element of the NOS not directly related to the VNOS questions and the number of answers that fell in each of coding categories is shown in Table XII. About the same number of BIO113 students mentioned at least one element of the NOS on the pre- and post-tests. The number of answers in each coding category increased or decreased slightly between the pre- and post-tests. Many more BIO114 students mentioned at least one element of the NOS on the post-test compared to the pre-test. In particular, the number of BIO114 answers in the coding categories "variables" and "scientific knowledge builds and/or changes" increased in the post-test compared to the pre-test. Interestingly, despite the overall increase in the number of BIO114 students who mentioned a specific element of NOS, there was a decrease in the categories "experiments need to be repeated" and "scientific method (term)."

		Total	Scientific methods (description, may not include term)	Scientific method (term)	Variables	Theories or assumptions	Need for repeated experiments	Science builds or changes
BIO113	pre-test	32	11	4	9	6	10	1
	post-test	31	4	3	14	5	3	8
BIO114	pre-test	58	14	5	25	12	9	4
	post-test	84	13	0	35	11	4	10

TABLE XII: A table showing how many BIO113 and BIO114 student answers were placed in a second coding category on the pre- and post-tests.

Though the use of specific scientific terminology was useful as an indication of previous scientific instruction, little could be concluded from most student answers about their views of the nature of science. For example, Korus (BIO114) answered the first question on the pre-test by writing, “A controlled test with an independent and dependent variable.” While this answer shows that s/he has learned the terms independent and dependent variable, the answer gives no insight into whether s/he knows the meaning of these terms. Conversely, on the pre-test Kitchell (BIO114) showed that s/he did not understand the meaning of the terms when s/he wrote, “A study conducted in which an independent variable is manipulated by a dependent variable” and Gourd (BIO114) that s/he at least partially understood the meaning of the terms when s/he wrote, “An experiment is a procedure designed to test a hypothesis usually involving measuring the effects of an independent variable on a dependent variable.”

Across all coding categories related to a specific element of the NOS, 9 of the 32 BIO113 student answers on the pre-test showed evidence of at least a partially correct understanding of scientific terms or ideas and six answers showed evidence of an incorrect understanding of scientific ideas or the meaning of scientific terms. On the post-test, 10 of the 31 BIO113 student answers showed evidence of at least a partially correct understanding of scientific terms or ideas and one answer evidence of an incorrect understanding. Twenty of the 58 BIO114 student answers on the pre-test showed at least a partially correct understanding of scientific terms or ideas and 11 answers an incorrect understanding. A similar pattern was observed in the BIO114 post-test answers with 28 of the 84 answers showing a partially correct understanding and six answers an incorrect understanding of scientific terms or ideas.

Student responsibility

On the questions about whether teachers or students were more responsible for certain learning tasks, BIO113 and BIO114 students distinguished that teachers were predominately responsible for two of the learning tasks given. Students did not score these questions significantly differently than the questions about tasks students were predominately responsible for on the pre-test (Z=-0.3797 and p-value=0.0704) but scored these questions significantly lower on the post-test (Z=-29.4672 and p-value<2.2e-16). This shows that students perceived teachers to be more responsible for these tasks by the end of the semester and provides evidence that they were reading the questions before responding. Students in BIO113 and BIO114 did not answer the student responsibility questions significantly differently on the pre-test (Z=1.7447 and p-value=0.08104) or post-test (Z=-1.0128 and p-value=0.3111). Within each course, students answered the questions significantly differently between the pre- and post-tests and effect sizes were large (BIO113: Z=-28.0314 and p-value<2.2e-6, and effect size=0.8538; BIO114: Z=-40.5068, p-value<2.2e-6, and effect size=0.8551). The means and standard deviations for BIO113 and BIO114 responses to the responsibility questions on the pre- and post-tests are shown in Table XIII. For all but one question about a task teachers are predominately responsible for (students getting clear guidelines for writing lab reports), the mean response to each question decreased between the pre- and post-tests. For the one question in which the mean increased, the mean only increased in BIO113. This shows that overall students perceived themselves to be less responsible and their instructors to be more responsible for all tasks assessed at the end of the semester compared to the beginning.

		Remembering information from lab manual	Taking notes in class	Getting clear guidelines for writing lab reports	Arranging meetings with the instructor	Staying awake during lab
BIO-113	pre	5.94±1.16	6.73±0.53	3.05±1.81	6.33±0.83	6.53±0.87
	post	5.71±1.29	6.57±0.58	3.27±1.91	6.21±0.95	6.18±1.23
BIO-114	pre	5.95±1.31	6.80±0.49	3.85±2.15	6.46±0.79	6.58±0.95
	post	5.27±1.34	6.42±0.80	3.34±1.80	6.20±0.94	6.18±1.14

		Learning the content covered in lab	Writing good lab reports	Participating in class discussions	Receiving timely feedback on their work	Reading assigned materials	Following the directions on assignments
BIO-113	pre	6.05±1.29	6.25±0.84	6.21±1.02	1.88±1.38	6.55±0.87	6.60±0.65
	post	6.03±1.20	5.99±0.99	5.88±1.13	1.78±1.39	6.12±1.05	6.16±0.96
BIO-114	pre	6.13±1.27	6.42±0.86	6.43±0.99	1.97±1.32	6.72±0.61	6.74±0.56
	post	5.66±1.25	5.92±0.92	5.64±1.04	1.82±1.33	6.11±1.11	6.16±0.97

TABLE XIII: Tables showing the mean response and standard deviation of BIO113 and BIO114 student responses to each responsibility question on the pre- and post-tests. Note that all means decrease from pre- to post-tests except for BIO113 students' mean response to "getting clear guidelines for writing lab reports."

In contrast to the student responsibility questions in which students rated teachers to be more responsible for learning and instructional tasks at the end of the semester, BIO114 student learning logs showed little change in content over the course of the semester. Eighteen out of the 20 students whose learning logs were coded made comments about how they prepared for assignments or planned to improve on future assignments in both the first and second set of learning logs. Within these broad categories of preparation and improvement, there was also little change within sub-categories, with most students discussing personal actions they took to prepare or would take to improve. One exception was the number of students who discussed ways to improve that included both personal actions and working with other people. Eleven people made comments that fell in this category on the first set of learning logs but only three of them did so on the second set. The number of student answers on learning logs that related to preparation or improvement in the first and second set of learning logs is shown in Table XIV.

		Total	Personal actions	Alone and with others	Only based on others	No related comments
Preparation	First half of semester	19	17	3	0	1
	Second half of semester	18	17	2	1	
	Made comments on both	18	15	2	0	
Improvement	First half of semester	18	18	11	2	0
	Second half of semester	19	19	5	3	
	Made comments on both	18	18	3	1	

TABLE XIV: Table XIV: A table showing how many students made comments that related to preparation and improvement in the first and second set of learning logs. The number of students who made comments on both sets of learning logs is also listed, along with the number of students who made no comments that fell into these categories on any of their eight learning logs.

The learning log answers in the preparation category that mentioned study methods were coded by whether or not the study methods mentioned were effective or ineffective. Almost all of the study methods students mentioned were ineffective, as shown in Table XV. In total, almost half of the students whose learning logs were coded mentioned having ineffective study methods on a learning log in the first and second half of the semester.

		Effective	Ineffective
Study methods	First half	1	12
	Second half	0	16
	Made comments on both	0	9

TABLE XV: A table showing how many students made comments about having effective or ineffective study methods on the first and second set of learning logs. The number of students who made comments on both sets of learning logs is also listed.

All 20 students whose learning logs were coded made comments about making mistakes. Eleven students explicitly mentioned making mistakes, and 15 implied they had made mistakes on both sets of learning logs. A few students made comments that suggested they were unsure if they made a mistake or gave reasons why they made a mistake on either the first and second set of learning logs, but only one student of the 20 whose learning logs were coded did so on both sets of learning logs. The number of student answers on learning logs that related to mistakes on the first and second set of learning logs is shown in Table XVI.

		Total	Explicitly mentioned	Implied	Unsure if they made a mistake	Made excuses	No related comments
	First half of semester	20	13	17	2	4	
Mistakes	Second half of semester	20	13	17	2	3	0
	Made comments on both	20	11	15	0	1	

TABLE XVI: A table showing how many students made comments that related to mistakes on the first and second set of learning logs. The number of students who made comments on both sets of learning logs is also listed, along with the number of students who made no comments that fell into these categories on any of their eight learning logs.

Eleven of the 20 students whose learning logs were coded made comments relating to grades, but only three did so on both the first and second set of learning logs. Whether students perceived the locus of control to be internal or external could not be determined for most answers. The number of student answers on learning logs that related to grades on the first and second set of learning logs is shown in Table XVII.

		Total	External locus of control	Internal locus of control	Mixed	No related comments
Grades	First half of semester	7	1	3	6	9
	Second half of semester	7	2	1	4	
	Made comments on both	3	0	0	2	

TABLE XVII: A table showing how many students made comments that related to grades on the first and second set of learning logs. The number of students who made comments on both sets of learning logs is also listed, along with the number of students who made no comments that fell in these categories on any of their eight learning logs.

Higher-order thinking skills

Students in BIO114 demonstrated different amounts of higher-order thinking skills (HOTS) depending on the question and assignment. There was no consistent trend across the semester. On quiz two, the students' first content quiz taken the second week of class, the mean performance for three questions that required HOTS were 3.03±1.58, 5.33±1.80, and 4.55±1.50 (the highest possible score was 9 for all HOTS assessments). On quiz six, taken on the seventh of 13 weeks of class, the mean performance was 5.29±1.40. On lab report one, which students submitted the tenth week of class, the mean performance was 4.60±1.24 and the mean performance on lab report two, submitted the final week of class, the mean performance was 4.67±1.59. The breakdown of the mean score for claim, evidence, and reasoning for each question is shown in Table XVIII.

Students' performance on most questions were significantly different from other questions, regardless of whether the questions were on the same assignment or the time in the semester when the assignment was completed. This is shown in Table XIX.

	Claim	Evidence	Reasoning
Quiz 2 Question 1	1.75±0.80	1.09±0.67	0.23±0.56
Quiz 2 Question 2	1.81±1.04	1.77±0.71	1.77±0.58
Quiz 2 Question 3	2.06±0.81	1.12±0.93	1.37±0.63
Quiz 6 Question 1	2.06±0.45	1.30±0.67	1.93±0.71
Lab Report 1	1.95±0.43	1.25±0.52	1.39±0.66
Lab Report 2	1.75±0.83	1.46±0.58	1.45±0.62

TABLE XVIII: The table above shows the mean and standard deviation of each question component (claim, evidence, and reasoning) on the different BIO114 quiz questions and lab reports that were analyzed to determine student development of higher-order thinking skills. 102 BIO114 student answers were analyzed.

	Quiz 2 Question 1	Quiz 2 Question 2	Quiz 2 Question 3	Quiz 6 Question 1	Lab Report 1
Quiz 2 Qu. 2	Z=-8.0573, p-value=7.803e-16				
Quiz 2 Qu. 3	Z=-6.9445, p-value=3.798e-12	Z=3.3149, p-value=0.0009168			
Quiz 6 Qu. 1	Z=-7.7327, p-value=1.0503e-14	Z=0.747, p-value=0.4551	Z=-3.4511, p-value=0.0005584		
Lab Rpt. 1	Z=-6.5721, p-value=4.961e-11	Z=4.2279, p-value=2.359e-5	Z=0.2611, p-value=0.794	Z=3.9227, p-value=8.758e-5	
Lab Rpt. 2	Z=-6.8033, p-value=1.023e-11	Z=3.2845, p-value=0.001022	Z=-0.2806, p-value=0.779	Z=3.1729, p-value=0.001509	Z=-0.2101, p-value=0.8336

TABLE XIX: The Z-scores and p-values associated with each comparison of a quiz question from the early to middle of the course and lab reports from the end of the of course. Notice that while many answers are significantly different, quiz two question three answers and the explanations given in the discussions of lab reports one and two are not significantly different from each other. The explanations given in lab report one are also not significantly different from those in lab report two.

DISCUSSION

The purpose of this study was to identify whether constructing an instructional plan to address student understanding of the nature of science, acceptance of responsibility for their work and performance in the course, and development of higher order thinking skills (frequently identified problems identified by biology teaching assistants at the institution where this study was conducted) would improve student performance in the targeted areas. It was predicted that the instructional plan would improve student performance in targeted areas more than traditional lab instruction, which was determined by comparing student performance in the lab in which the study was implemented to another introductory biology lab. Overall, student performance did not improve as expected, although some small changes in student performance where observed.

Student-based needs assessment survey

There were differences in how frequently instructional challenges or gaps were reported in the first and second versions of the student-based needs assessment survey (SB-NAS). A likely reason for this is that the results of version two were more representative of the different science departments. Version one of the survey was taken by predominately biology TAs, with 17 of the 24 TAs who took the survey being TAs in the biology department. The second version was taken by more TAs in the chemistry and physics departments and fewer biology TAs, with TAs in the biology department making up 13 of the 26 TAs who completed the survey. Therefore, TAs in the chemistry and physics department had a larger impact on the most frequent gaps identified by SB-NAS ver.2.

Further evidence supporting the explanation that the results of SB-NAS version 2 better represented those experienced by TAs in all science departments can be found by looking at the gaps identified as most frequent in SB-NAS versions one and two. SB-NAS ver.1 identified four frequent challenges: 1) student organization of biology knowledge, 2) student development of higher-order thinking skills, 3) student understanding of the nature of science, and 4) student acceptance of responsibility for their work and performance in the course. SB-NAS ver.2

identified three of the same gaps as most frequent, but student acceptance of responsibility (identified by version 1) was replaced by student ability to work effectively in groups in version two. This can be explained by looking at how frequently each gap was reported to be experienced by TAs in different science departments on SB-NAS ver.2 (Appendix E), student responsibility was scored lower by TAs in the chemistry and physics departments (40% and 17% of responses were frequent or routine, respectively) compared to TAs in the biology department who selected students accepting responsibility to be a frequent or routine problem 45% of the time. This could explain why student responsibility was not identified as one of the most frequent gaps in SB-NAS version two like it was in version one.

In contrast to student responsibility, students not working effectively in groups was reported to be a more frequent problem in SB-NAS ver.2 compared to ver.1. This can also be explained by looking at how frequently this gap was reported to be experienced by TAs in different science departments on SB-NAS ver.2 (Appendix E). Students not working effectively in groups was reported to be frequent or routine a larger percentage of the time by TAs in chemistry (60%) and physics (67%) compared to TAs in the biology department (58%). This could explain why working in groups was not identified as one of the most frequent gaps in SB-NAS ver.1, which was predominately taken by biology TAs, compared to SB-NAS ver.2.

SB-NAS ver.2 may have also been more representative of TAs in all science departments because it contained questions that were more applicable to TAs in all science departments. As shown in Appendix B, the scenarios used in SB-NAS ver.2 contained little to no biology content, compared to the scenarios used in SB-NAS ver.1 (Appendix A). In addition to removing content, minor adjustments in wording were made to account for differences in the format of biology, chemistry and physics departments at the university where the survey was administered. These changes may have changed the perceived relevancy of different scenarios to chemistry and physics TAs, thus changing the most frequently identified gaps in SB-NAS ver.2. In this way, both the composition of SB-NAS ver.2 and the number of TAs in each science department who completed SB-NAS ver.2 resulted in this version of the survey being a better reflection of the gaps experienced by TAs in biology, chemistry and physics departments.

The results of SB-NAS vers.2 were more representative of different science departments and the survey's internal reliability, as evidence by McDonald's omega, was higher than that of SB-NAS ver.1. Future iterations of this study should, therefore, use SB-NAS ver.2 to identify the most frequent gaps experienced by science TAs.

Overall findings

The revised instructional plan implemented in BIO114 did not produce the hoped for increases in student understandings of the nature of science, student acceptance of responsibility, or higher-order thinking skills. Explanations for why there were no large increases in each gap are discussed in the following sections.

Nature of Science

Though student views of the NOS did not change for all elements assessed by the VOSTS question set as a whole, BIO114 students showed evidence of a growing understanding of scientific knowledge as fluid compared to BIO113 students. One of the VOSTS questions, 90411, asked students their position on whether scientific knowledge can change. Answer choice A, the appropriate answer choice, was the only answer that stated scientific theories or discoveries can be disproven and thus change. Eleven percent fewer BIO113 students selected this answer choice on the post-test compared to the pre-test. Conversely, 8% more BIO114 students selected this answer choice on the post-test compared to the pre-test. Though the sample size is small, the fact that BIO113 and BIO114 students show opposite trends in understanding scientific knowledge can change, suggests that the BIO114 curriculum improved students' understanding that scientific knowledge is fluid while the BIO113 curriculum did not.

Further support for student views of the NOS becoming more informed as a result of the BIO114 curriculum can be found from the VNOS questions. Student answers to these questions were categorized by whether or not students used language that implied they thought scientific knowledge was absolute or non-absolute. Students who used language that implied they thought scientific knowledge was not absolute were more likely to understand that scientific knowledge was evidence-based but capable of changing based on new findings. For VNOS question one, 50% fewer BIO113 students used non-absolute language when referring to scientific knowledge and 5% fewer BIO113 students used less absolute language when referring to scientific knowledge on the post-test compared to the pre-test. The decrease in both answer types was coupled with an increase (33%) in the number of answers that used vague answers that could not be classified. Due to the fact both types of answers (absolute and non-absolute) decreased between the pre- and post-tests, there is no evidence to support BIO113 student views of science as fluid as becoming more or even less informed.

In contrast to BIO113 students, 25% more BIO114 student answers to VNOS question one used non-absolute language when referring to scientific knowledge on the post-test compared to the pre-test and roughly 43% less BIO114 students used absolute language. This provides evidence that the BIO114 curriculum helped students understand that scientific knowledge can change. BIO114 student answers to VNOS question two also support this conclusion, as 51% fewer students used absolute language and 130% more students used non-absolute language to refer to scientific knowledge on the post-test compared to the pre-test. Unlike with VNOS question one, this trend was also seen in BIO113 student answers, but the changes were much smaller than in BIO114 (11% decrease in absolute language and 7% increase in non-absolute language). While this could be an indication that BIO113 students' understanding of science as a fluid field also improved, the changes in language used were very small, the result of only 1-2 students removing absolute or adding non-absolute language, compared to VNOS question one where many more students changed their answer types between the pre- and post-tests. For this reason, it does not appear that BIO113 students' views had changed as a result of taking the course.

There is evidence to suggest that in addition to coming to understand that scientific knowledge can change, BIO114 student views of other elements of the NOS changed as a result of taking the course. In addition to categorizing VNOS answers as using absolute or non-absolute language when referring to scientific knowledge, answers that mentioned scientific terms or ideas other than those directly assessed by the questions were coded. BIO113 students mentioned other scientific terms or ideas roughly the same number of times (there was a 3% decrease) on the pre- and post-test. Conversely, in BIO114 there was a 45% increase in the number of answers that mentioned other scientific terms or ideas. This suggests that students' understanding of other aspects of the NOS had changed as a result of completing BIO114 but not BIO113. However, more research is needed to support this claim. In particular, explicit assessment of student understandings of other areas of the nature of science targeted by the BIO114 curriculum, including science as question driven and evidence-based, should be added. Little can be concluded about these elements from this study because they were not directly assessed.

In future iterations of this study, more care should be taken in making sure that the instruction on the NOS students receive in class matches the NOS assessments. This could be accomplished by adding or removing VOSTS questions to ensure that they assess selected elements, or even by writing similarly formatted questions if none currently exist. Such changes would allow for a more quantitative measure of whether student views of the NOS are becoming more informed as a result of the BIO114 curriculum. Future studies will also increase sample size, allowing for stronger conclusions to be made about student views based on the VNOS questions, which is currently limited by the small number of answers that fall in each coding category.

Student responsibility

Student acceptance of responsibility for specific learning tasks assessed by the questions asked on the pre- and post-tests (Appendix D), decreased between the pre- and post-tests in both BIO113 and BIO114 students. No difference was seen in BIO114 students accepting responsibility on learning log answers, possibly due to small sample size. This decrease in perceived responsibility can be attributed to the low-quality study methods of the students. As previously described, Zimmerman and Kitsantas (2005) found low-quality study methods to be indirectly linked to accepting less responsibility for learning tasks and outcomes because engaging in low-quality study methods lowered the students' self-efficacy for learning. This explanation is supported by the comments BIO114 students made about their study methods on learning logs. All study methods students described but one were ineffective and 9 of the 20 students whose learning logs were analyzed described using ineffective study methods (passive actions like rereading or highlighting) in a learning log completed in the first and second half of the semester.

In future iterations of the instructional plan, effective study methods should be explicitly discussed with students in more class periods. Discussing effective study methods the first class period and reminding students of this discussion in personal meetings, as was done in the revised instructional plan, was not sufficient to change students' study habits. In addition to adding more instruction on how to study, a question about study methods should be added to the learning log instruction sheet so that the study methods of all students can be analyzed. This will allow for stronger conclusions to be made about students' study methods, as this study's conclusions are limited by the lack of direct assessment of students' study methods.

Higher-order thinking skills

The explanations students provided on quizzes and lab reports throughout the semester provided no evidence of students' higher-order thinking skills (HOTS) improving as the semester progressed. In fact, the HOTS students demonstrated varied by question, regardless of what assignment the question was a part of or when in the semester the assignment was completed. This could be because the HOTS students demonstrated varied based on the level of support provided and not on the actual skills students possessed. Students demonstrated the most HOTS on two quiz questions, one on quiz two and one on quiz six. This is likely because most or all TAs discussed both of these questions in the class session immediately preceding the quiz. The question on quiz two asked students about the main findings of the previous weeks' lab and thus there is a high probability it was discussed in class and group discussions. The question on quiz six was a review question from the first unit (the same unit as the quiz two questions). Because quiz six was a review quiz, all the information on this quiz was reviewed in an in-class activity.

The two quiz questions on quiz two with lower means and the students' explanations for their data given in their lab reports may not have been explicitly discussed in all sections of BIO114. In the initial TA training, TAs were briefly instructed in how to guide student thinking instead of giving them answers and in types of questions that required higher levels of thinking. In addition, TAs were given all quiz questions requiring higher-order thinking well in advance of when the quiz was administered so that they could prepare students for these questions. Whether TAs used this information and asked their students similar questions in class is unknown. A previous study found that teachers who went through 40 hours of training asked significantly more questions requiring higher-order thinking than other teachers (Gul et al., 2014). Since TAs in this study went through significantly less training, it is reasonable to speculate that TAs in this study might have only asked students higher-order thinking questions that were directly incorporated into in-class activities or other course assignments. Alternatively, different TAs might have spent different amounts of time developing HOTS. Clearly this is an issue to explore in more depth as part of any future studies. For example, could a short daily TA debriefing of each class session help identify possible covariates? Do TAs need more explicit instructions or follow-up training on pedagogical approaches to HOTS development?

The differences in HOTS demonstrated on quiz questions and student explanations not discussed in class can be attributed to each assignment's directions. Quizzes included a generic set of directions that asked students to support all explanations with evidence and explain their reasoning. However, some students may not have known what counted as proper evidence or reasoning and thus, may have demonstrated less HOTS than they would have had they been instructed what evidence or reasoning to provide. Conversely, lab report instructions told students to give a physiological explanation for their results and to provide evidence supporting this explanation. They also told students that part of this evidence should come from comparing their explanation to the results and interpretations of previously published studies. In this way, students were given more assistance in deciding what evidence and reasoning to include on lab reports than they were on quizzes. This could explain why BIO114 students demonstrated increased HOTS on lab reports compared to quizzes, even though both asked students to develop explanations. These results suggest another question for future study: would students benefit from more directed, explicit training and scaffolding on rules of evidence and reasoning?

In future iterations of this study, care should be taken to ensure that students in lab sections taught by all TAs are adequately prepared for HOTS assessments. This can be accomplished by having TAs develop and discuss higher-order thinking questions they can ask in class and by ensuring that all HOTS assessments tell students what reasoning and evidence are necessary to support their claims. TAs would likely benefit from more explicit training on HOTS question writing as well. This will also allow for conclusions to be made about how students' HOTS are developing over the course of the semester.

Additional thoughts

An alternative explanation for why only small changes were observed in students' understanding of the NOS, student acceptance of responsibility for their learning decreased, and demonstrated HOTS skills varied throughout the semester is that the students' developmental stage hindered larger changes. As shown in Tables I and II, roughly two-thirds of students enrolled in both BIO113 and BIO114 in Fall 2015 were first- or second- year students. It has been found that eighty-six percent of first-year students and fifty-seven percent of second-year students solely depend on external authorities (Magolda, King, Taylor, & Wakefield, 2012). This means that the majority of first- and second-year students accept no responsibility for their work and performance and depend on their instructors to tell them what to do and when to do it. This could explain why BIO114 students did not see themselves as learners as being more responsible for certain learning tasks even though this was emphasized in the curriculum.

It has also been proposed that students' intellectual development proceeds through a set of stages, of which the first is seeing all knowledge as right or wrong, the second is that knowledge is merely an opinion, and the third sees knowledge as not absolute but still based on evidence (Perry, 1968). This could explain why students' answers did not show an understanding of the concept that scientific knowledge can change and why they struggled to give plausible explanations supported by evidence and biological principals. If this is the case, then the impact of the instructional plan may thus not be seen until later in their college years when they have reached a later stage of intellectual development. More research is necessary to explore how students' developmental stage may be impacting their performance.

Conclusion

This study showed that student understandings of the nature of science and acceptance of responsibility for their work and performance change over the course of taking a biology lab. Similar conclusions could not be made about students' higher-order thinking skills due to differences in how much assignments requiring higher-order thinking skills were discussed in class. Changes should be made to the BIO114 curriculum to include more explicit instruction on key elements of the NOS, effective study methods, and the HOTS assessments used in the course. It is expected that these changes will cause students to experience greater gains in understanding key elements of the nature of science, accepting responsibility for their work and performance in the course, and in developing higher-order thinking skills.

WORKS CITED

Aikenhead, G. S., Ryan, A. G., & Fleming, R. W. (1989). Views on Science-Technology-Society, Form CDN.mc.5. Dept. Curriculum Studies, College of Education, University of Saskatchewan, CA. Retrieved 1/1/15, https://www.usask.ca/education/profiles/aikenhead/webpage/vosts.pdf

Anderson, L. W., & Krathwohl, D. R., editors (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, 1st edition. New York, NY: Pearson, Inc. 336pp.

Anderson, T. R., & Rogan, J. M. (2011). Bridging the educational research-teaching practice gap: Curriculum development, Part 1: Components of the curriculum and influences on the process of curriculum design. *Biochemistry and Molecular Biology Education*, 39(1), 68–76. DOI: 10.1002/bmb.20470

Angelo, T. A., & Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*, 2nd edition. San Francisco, CA: Jossey-Bass, Inc. 448pp.

Association of American Colleges and Universities. (2005). *Liberal Education Outcomes: A Preliminary Report on Student Achievement in College*. Retrieved February 22, 2016, from http://www.aacu.org/sites/default/files/files/LEAP/LEAP_Report_2005.pdf

Barry, D., & Dotger, B. (2011). Enhancing content knowledge in graduate teaching assistants through lesson study. Paper presented at the 2011 National Conference on Research Based Undergraduate Science Teaching. Tuscaloosa, AL.

Baumgartner, E. (2007). A professional development teaching course for science graduate students. *Journal of College Science Teaching*, 36(6), 16–21.

Brown, P.C., III, Roediger, H.L., McDaniel, M.A. (2014). *Make It Stick: The Science of Successful Learning*. Cambridge, Massachusetts: Belknap Press. 336 pp.

Commission on the Future of Higher Education. (2006). *A Test of Leadership: Charting the Future of U.S. Higher Education, Pren-Publication Copy* -- September 2006 (PDF) - pre-pub-report.pdf. U.S. Department of Education.

Committee on How People Learn, National Research Council of the National Academies, & Donovan, M.S., and Bransford, J.D., editors. (2005). *How Students Learn: Science in the Classroom*. Washington, D.C.: National Academies Press. 264pp.

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction.* Research Findings and Recommendations. Research Findings and Recommendations Prepared for the Committee on Pre-College Philosophy of the American Philosophical Association, Newark, DE. 110pp. Retrieved 1/1/15 from http://files.eric.ed.gov/fulltext/ED315423.pdf

Fishman, E. J. (2014). With great control comes great responsibility: the relationship between perceived academic control, student responsibility, and self-regulation. *The British Journal of Educational Psychology*, 84, 685–702. DOI: 10.1111/bjep.12057

Gardner, G. E., & Jones, M. G. (2011). Pedagogical preparation of the science graduate teaching assistant: Challenges and implications. *Science Educator*, 20(2), 31–41.

Golde, C. M., & Dore, T. M. (2001). *At cross purposes: what the experiences of doctoral students reveal about doctoral education*. A survey initiated by Pew Charitable Trusts. Philadelphia, PA: Pew Charitable Trusts. 58pp. Retrieved 1/1/16 from http://www.phdcompletion.org/promising/Golde.pdf

Gul, R. B., Khan, S., Ahmed, A., et al. (2014). Enhancing educators' skills for promoting critical thinking in their classroom discourses: A randomized control trial. *International Journal of Teaching and Learning in Higher Education*, 26(1), 37–54.

Gupta, T., Burke, K. A., Mehta, A., & Greenbowe, T. J. (2015). Impact of Guided-Inquiry-Based Instruction with a Writing and Reflection Emphasis on Chemistry Students' Critical Thinking Abilities. *Journal of Chemical Education*, 92(1), 32–38. DOI: 10.1021/ed500059r

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449–455. DOI: 10.1037/0003-066X.53.4.449

Handelsman, J., Miller, S., & Pfund, C. (2006). *Scientific Teaching*, 1st edition. The Wisconsin Program for Scientific Teaching, supported by the Howard Medical Institute Professors Program. New York, NY: W. H. Freeman and Company. 184pp.

Hart Research Associates. (2013). *It takes more than a major: Employer priorities for college learning and student success.* An online survey among employers conducted on behalf of: The Association of American Colleges and Universities. Washington, DC.

Hothorn, T., Hornik, K., van de Wiel, M.A., Winell, H., Zeileis, A. (2015). *Conditional inference procedures in a permutation test framework.* Coin package (R version 3.1.3). Retrieved April 6, 2016 from https://cran.r-project.org/web/packages/coin/coin.pdf

How U.S. News Calculated the 2015 Best Colleges Rankings - US News. (n.d.). Retrieved May 27, 2015, from http://www.usnews.com/education/best-colleges/articles/2014/09/08/how-us-news-calculated-the-2015-best-colleges-rankings

Hugerat, M., & Kortam, N. (2014). Improving higher-order thinking skills among freshmen by teaching science through inquiry. *EURASIA Journal of Mathematics, Science & Technology Education*, 10(5), 447–454.

Kelly, K., & Lai, K. (2012). MBESS package (R version 3.3.3). Retrieved from http://nd.edu/~kkelley/site/MBESS.html

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

Kim, K., Sharma, P., Land, S. M., & Furlong, K. P. (2013). Effects of active learning on enhancing student critical thinking in an undergraduate general science course. *Innovative Higher Education,* 38(3), 223–235. DOI: 10.1007/s10755-012-9236-x

Lauermann, F., & Karabenick, S. A. (2011). Taking teacher responsibility into account(ability): Explicating its multiple components and theoretical status. *Educational Psychologist,* 46(2), 122–140. DOI: 10.1080/00461520.2011.558818

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

Lederman, N. G., & O'Malley, M. (1990). Students' perceptions of tentativeness in science: Development, use, and sources of change. *Science Education*, 74(2), 225–239. DOI: 10.1002/sce.3730740207

Lederman, N. G., & Zeidler, D. L. (1987). Science teachers' conceptions of the nature of science: Do they really influence teaching behavior? *Science Education*, 71(5), 721–734. DOI: 10.1002/sce.3730710509

Luft, J. A., Kurdziel, J. P., Roehrig, G. H., & Turner, J. (2004). Growing a garden without water: Graduate teaching assistants in introductory science laboratories at a doctoral/research university. *Journal of Research in Science Teaching*, 41(3), 211–233. DOI: 10.1002/tea.20004

Magolda, M. B. B., King, P. M., Taylor, K. B., & Wakefield, K. M. (2012). Decreasing authority dependence during the first year of college. *Journal of College Student Development,* 53(3), 418–435. DOI: 10.1353/csd.2012.0040

McNeill, K. L., & Krajcik, J. S. (2011). *Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing*, 1st edition. Boston, MA: Pearson, Inc. 208pp.

National Research Council (U.S.), Donovan, M.S., & Brandford, J.D. (2005). *How Students Learn: History, Mathematics, and Science in the Classroom.* Washington, DC: National Academies Press. 264pp.

National Survey of Student Engagement. (2013). *A Fresh Look at Student Engagement- Annual results 2013*. Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved 1/1/16 from http://nsse.indiana.edu/nsse_2013_results/pdf/nsse_2013_annual_results.pdf

National Survey of Student Engagement. (2010). *Major differences: Examining student engagement by field of study- Annual results 2010.* Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved 1/1/16 from http://nsse.indiana.edu/nsse_2010_results/pdf/nsse_2010_annualresults.pdf

Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15*. Milton Keynes, UK, USA: Open University Press. pp225.

Pascarella, E. T., & Blaich, C. (2013). Lessons from the Wabash National Study of Liberal Arts Education. *Change: The Magazine of Higher Learning*, 45(2), 6–15. DOI: 10.1080/00091383.2013.764257

Perry, W. G. (1968). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York, NY: Holt, Rinehart & Winston. 336 pp.

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Smith, C. L., & Wenk, L. (2006). Relations among three aspects of first-year college students' epistemologies of science. *Journal of Research in Science Teaching*, 43(8), 747–785. DOI: 10.1002/tea.20113

Smith, D. G. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology*, 69(2), 180–190. DOI: 10.1037/0022-0663.69.2.180

Smith, V. G., & Szymanski, A. (2013). Critical thinking: more than test scores. International *Journal of Educational Leadership Preparation*, 8(2), 16–25.

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE - Life Sciences Education*, 11(3), 294–306. DOI: 10.1187/cbe.11-11-0100

Stein, M., Larrabee, T. G., & Barman, C. R. (2008). A study of common beliefs and misconceptions in physical science. *Journal of Elementary Science Education*, 20(2), 1–11.

Sundberg, M. D., Armstrong, J. E., & Wischusen, E. W. (2005). A reappraisal of the status of introductory biology laboratory education in U.S. colleges & universities. *American Biology Teacher,* 67(9), 525–529.

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *The Journal of Higher Education*, 73(6), 740–763. DOI: 10.1353/jhe.2002.0056

Understanding Science. (2016). University of California Museum of Paleontology. Retrieved February 2, 2016, from http://undsci.berkeley.edu/

Vazquez-Alonso, A., & Manassero-Mas, M.-A. (1999). Response and scoring models for the "Views on Science-Technology-Society" instrument. *International Journal of Science Education*, 21(3), 231–247. DOI: 10.1080/095006999290679

Wenk, L., & Tronsky, L. (2011). First-year students benefit from reading primary research articles. *Journal of College Science Teaching*, 40(4), 60–67.

Wiggins, G., & McTighe, J. (1998). *Understanding by Design,* 1st edition. Alexandria, VA: Pearson Education. 201pp.

Zimmerman, B. J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology*, 30(4), 397–417. DOI: 10.1016/j.cedpsych.2005.05.003

APPENDIX A

Student based-needs assessment survey (SB-NAS) version 1

1. What university are you attending?
   - ☐ Wake Forest University

2. What department are you in?
   - ☐ Chemistry
   - ☐ Biology
   - ☐ Physics

3. Are you a master's or PhD student?
   - ☐ Master's
   - ☐ PhD

4. How many semesters have you been a teaching assistant (TA)?
   - ☐ 0
   - ☐ 1
   - ☐ 2
   - ☐ 3
   - ☐ 4
   - ☐ 5
   - ☐ 6
   - ☐ 7
   - ☐ 8
   - ☐ 9
   - ☐ 10
   - ☐ 11
   - ☐ 12

5. Have you been a TA continuously?
   - ☐ Yes
   - ☐ No

6. What has been your longest break in teaching?
   - ☐ One semester
   - ☐ One year
   - ☐ A year and a half
   - ☐ 2 years
   - ☐ 2 and a half years
   - ☐ 3 years
   - ☐ More than 3 years

7. How many semesters have you been a graduate student?
   - □ 1
   - □ 2
   - □ 3
   - □ 4
   - □ 5
   - □ 6
   - □ 7
   - □ 8
   - □ 9
   - □ 10
   - □ 11
   - □ 12
   - □ More than 12

8. Did you have any teaching experience before becoming a TA in your current position?
   - □ Yes
   - □ No

9. Do you plan on becoming a teacher or professor?
   - □ Yes
   - □ No

Student self-efficacy

1. A student comes into their TA's office to talk about a lab report s/he handed back the day before in class. The student earned a C- on the assignment. The student asks a few questions about what s/he did wrong, then starts crying and says s/he will never be able to do well on a lab report. The TA tries to encourage the student that this is the first of four lab reports and that if s/he works hard s/he will improve. The student leaves the meeting still crying and ends up earning a C on all four of his/her lab reports. As a TA, how often have you experienced something similar to this?
   - □ Never
   - □ Once a semester
   - □ Two times a semester
   - □ Three or more times a semester
   - □ Almost every week

2. Mid-semester, a student meets with his/her TA to talk about how class is going. The student has been doing well and has a B average in lab. At the meeting, the student asks questions about his/her grades on a few different assignments. The next week in lab the TA overhears him/her telling a classmate this will be his/her last course in biology. When the TA asks why, s/he remarks that biology is the only class s/he does not currently have an A in, so s/he is obviously not meant to be a biology major. As a TA, how often have you experienced something similar to this?
   - □ Never
   - □ Once a semester
   - □ Two times a semester
   - □ Three or more times a semester
   - □ Almost every week

Student motivation/ strategic learning

1. A TA is assigned to teach a course in which students conduct a series of related lab exercises all semester, then make and present a poster on their project. At the beginning of each lab, the TA encourages students to write down their methods, results and observations. Two weeks later, students still are not following these instructions. Frustrated, the TA tells students s/he is going to start checking their notebooks every week. A student raises his/her hand and asks if their notes will be graded. The TA says no but taking notes will help them make their final poster. The next week, only 2 of the 20 students in the class have started making notes. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

2. It's nearing the end of the semester and a TA has just finished grading and commenting on students' third lab reports in a second-year course. The TA is disappointed that the majority of students have not improved from the first paper, so the TA decides to let the students rewrite their papers. After handing back the papers in lab, the TA informs the students they can rewrite their papers. A student raises his/her hand and asks if fixing errors that the TA commented on are all s/he needs to do to get an A on the rewrite. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Knowledge organizations/ the big picture

1. A TA is starting a new unit in lab on plant function. At the beginning of lab, the TA gives a brief lecture on transpiration explains how plants must balance photosynthesis and water loss. Students then complete a lab investigating how much plants transpire at different levels of sun. For homework, the TA asks students to predict how much plants would photosynthesize if given different amounts of water. Only 4 of 20 students in class give a logical answer that incorporates facts learned in lab. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

2. In a first-year lab course, students write a lab report on an experiment they design. Students are expected to give relevant background information, describe their methods, and summarize and provide an explanation for their results. When grading students' lab reports, a TA finds that students fail to connect ideas presented in different parts of their paper. For example, students discuss information in their introduction that they learned in class that is not relevant to their particular experiment. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week
- □ Not applicable

Knowledge and skill transfer

1. A student works hard and gets an A on their second lab report. Their TA is particularly impressed by the logical explanation the student gives for their results. On the third lab report, the student does not perform as well and his/her explanation for his/her results makes no sense. The TA asks the student why s/he thinks s/he did not do as well on his/her third paper. She responds that s/he was not sure how to write the different parts of the paper or explain his/her results for the last experiment because it was so different from his/her second experiment. As a TA, how often have you experienced something similar to this?
    - □ Never
    - □ Once a semester
    - □ Two times a semester
    - □ Three or more times a semester
    - □ Almost every week

2. At the beginning of the semester, students dissect crayfish to learn about their respiratory systems. A few weeks later, a group of students designs an experiment to study effects of exercise on crayfish metabolism. They propose putting the crayfish on the lab bench and chasing it with a pencil. When their TA asks them how the crayfish will be able to breathe, they answer that it will not, and propose exercising it in a tub of water. Their TA explains that this is not necessary because crayfish can breathe as long as its gills are wet. When the TA asks the students to explain why, none of them answers. As a TA, how often have you experienced something similar to this?
    - □ Never
    - □ Once a semester
    - □ Two times a semester
    - □ Three or more times a semester
    - □ Almost every week

Students accepting responsibility for their work

1. A student turns in a homework assignment with several factually wrong answers. The assignment also has multiple grammatical and spelling mistakes. In all, the assignment receives a score of 60%. When the TA hands back the assignment at the end of class, the student takes one look at his/her grade and requests a meeting with his/her TA, saying that s/he does not understand their expectations. As a TA, how often have you experienced something similar to this?
    - □ Never
    - □ Once a semester
    - □ Two times a semester
    - □ Three or more times a semester
    - □ Almost every week

2. Students in a lab turn in their first reports describing their own experiments. Many of the reports are incomplete, are missing many key parts, and contain typos and writing errors. When the TA hands back papers, one student asks what the class average was. When the TA answers, several students proclaim that the TA must be grading too hard. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Higher-order thinking

1. A group of students has finished an experiment they designed investigating the effect of caffeine on muscle contraction. Their TA asks the students what they found. One student answers that they predicted that caffeine would increase the amplitude of muscle contraction but that their results showed caffeine had no effect on muscle contraction. The TA asks the students to think of reasons why their results could have been different from past experiments. None of the students attempt to answer. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

2. On every quiz and homework assignment a TA asks one question that requires students to apply facts and concepts learned in class to a new situation. Some questions ask about stories in the news, others ask students to think of new examples of concepts. Students frequently complain that these questions are too hard. At the end of the semester, the TA finds that most students, including students who get all other questions correct, consistently missed these questions. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Nature of Science

1. A TA is listening in on a group of students discussing what errors they could have made performing their experiment. The students are confident that they have made a mistake, they are just unsure what they did wrong. The TA asks them why they think they have made a mistake. The students reply that their results must be wrong because they do not match the prediction they made in their hypothesis. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

2. In a lab on hormones, a group of students decides to repeat a classical experiment looking at the role of juvenile hormone. Their results show that juvenile hormone does not prevent pupation, which contradicts previous experiments. The group goes to their TA and explains that their experiment did not work because their results do not match past studies. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Group work/ cooperative learning

1. A TA breaks up the class of 16 into 4 groups. Each week, groups work together to complete a given experiment. The TA explains that the groups can divide up the work but that they need to discuss the experiment together to make sure they all understand the material and know how to do every part. The third week of lab, the TA asks students to write a hypothesis and make a figure (with data provided) on a quiz. Most students miss at least one of the questions and write that another member of their group always writes their hypothesis or makes their graphs. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

2. At the end of the semester, a TA instructs groups of 4 students to pick a primary literature article and present it to their classmates. To assess their understanding of the article, the TA gives the groups a short assignment in class that includes questions on each section of the article. S/he instructs the students to complete the assignment as individuals. When grading the assignment, the TA notices that most of the students correctly answer questions about 1-2 of the sections of the article, but seem to know very little about the other sections of the article. As a TA, how often have you experienced something similar to this?

- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

APPENDIX B

Student Based-Needs Assessment Survey version 2

1. What department are you in?
   - ◻ Chemistry
   - ◻ Biology
   - ◻ Physics

2. Are you a master's or PhD student?
   - ◻ Master's
   - ◻ PhD

3. How many semesters have you been a teaching assistant (TA)?
   - ◻ 0
   - ◻ 1
   - ◻ 2
   - ◻ 3
   - ◻ 4
   - ◻ 5
   - ◻ 6
   - ◻ 7
   - ◻ 8
   - ◻ 9
   - ◻ 10
   - ◻ 11
   - ◻ 12
   - ◻ More than 12

4. Have you had any gaps or breaks in teaching? By this we mean you had a fall or spring semester in which you did not teach, between semesters when you did teach.
   - ◻ Yes
   - ◻ No

Survey branches based on response:

If **Yes** is selected, goes to "5. What has been your longest break in teaching?"

If **No** is selected, skips to "6. How many semesters have you been a graduate student?"

5. What has been your longest break in teaching?
   - ◻ One semester
   - ◻ One year
   - ◻ A year and a half
   - ◻ 2 years
   - ◻ 2 and a half years
   - ◻ 3 years
   - ◻ More than 3 years

6. How many semesters have you been a graduate student?
   - □ 1
   - □ 2
   - □ 3
   - □ 4
   - □ 5
   - □ 6
   - □ 7
   - □ 8
   - □ 9
   - □ 10
   - □ 11
   - □ 12
   - □ More than 12

7. Did you have any teaching experience before becoming a TA in your current position?
   - □ Yes
   - □ No

8. Do you plan on becoming a teacher or professor?
   - □ Yes
   - □ No

Student self-efficacy

Question 1: Generic version

A student comes into their TA's office to talk about their first lab report for which they earned a C-. The student asks a few questions about what they did wrong, then becomes emotional and says they will never be able to do well on a lab report. The TA encourages the student, saying that scientific writing is hard but if they work hard they will improve. The student leaves the meeting and continues earning Cs on subsequent lab reports.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Biology version

Mid-semester, a student meets with the TA to talk about how class is going. The student has been doing well and has a B average in lab. At the meeting, the student asks questions about their grades on a few different assignments. The next week in lab the TA overhears the student telling a classmate this will be their last course in biology. When the TA asks why, the student remarks that biology is the only class they do not currently have an A in, so they are obviously not meant to be a biology major.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Chemistry version

Mid-semester, a student meets with the TA to talk about how class is going. The student has been doing well and has a B average in lab. At the meeting, the student asks questions about their grades on a few different assignments. The next week in lab the TA overhears the student telling a classmate this will be their last course in chemistry. When the TA asks why, the student remarks that chemistry is the only class they do not currently have an A in, so they are obviously not meant to be a chemistry major.

Question 2: Physics version

Mid-semester, a student meets with the TA to talk about how class is going. The student has been doing well and has a B average in lab. At the meeting, the student asks questions about their grades on a few different assignments. The next week in lab the TA overhears the student telling a classmate this will be their last course in physics. When the TA asks why, the student remarks that physics is the only class they do not currently have an A in, so they are obviously not meant to be a physics major.

Student motivation/ strategic learning

Question 1: Generic version

At the beginning of each lab, a TA encourages their students to write down lab methods, results and observations because they will need to know this information when writing about the experiments. Two weeks later, students are still not following these instructions. The TA asks students why there are not taking notes. One student raises their hand and responds that they do not take notes because it does not count for a grade.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Generic version

A TA realizes that a few groups of students in the lab are constantly asking questions. In response, the TA starts asking several groups questions to see what parts of experiments they are failing to understand. The TA finds that the members of these groups are unable to answer basic questions about the purpose of the experiment and/or what they are doing in the experiment.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Knowledge organizations/ the big picture

Question 1: Generic version

Students complete several exercises and experiments during a unit that spans multiple weeks. At the end of the unit, students are asked questions that require them to synthesize information learned throughout the unit. The TA notices that some students who can correctly answer questions about the different parts of the unit, are not able to answer these synthesis questions.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Generic version

In the conclusion/discussion section of a lab report a TA asks students to summarize their experiments. When looking over this section of students' lab reports, the TA notices that some students fail to connect ideas presented in different parts of their lab reports.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Knowledge and skill transfer

Question 1: Generic version

A TA is impressed by how well one student explains the results of an experiment. However, on the next lab report the student gives an illogical interpretation of her results. The student arranges a meeting with the TA and shares that they struggled with interpreting her results because the last experiment was so different from the other experiments completed in lab.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Generic version

During a lab the first few weeks of class, a TA teaches students how to conduct a basic lab procedure. Most students successfully carry out this procedure the following week. A month later, students must conduct the same procedure as a part of a different experiment. Some students struggle to conduct the procedure despite having written instructions.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Students accepting responsibility for their work

Question 1: Generic version

A student turns in a homework assignment with several factually wrong answers. It also has multiple grammatical and spelling mistakes. In all, the assignment receives a score of 60%. When the TA hands back the assignment at the end of class, the student takes one look at her grade and requests a meeting with her TA, saying that they does not understand their expectations.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Generic version

Students turn in their first lab reports. The TA finds that many are incomplete, missing many key parts, and contain typos and writing errors. When the TA hands back papers, one student asks what the class average was. When the TA answers, several students proclaim that the TA must be grading too hard.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Higher-order thinking

Question 1: Biology version

A group of students has just finished an experiment. They find that their results are the opposite of what they expected. Knowing students are required to explain their results in lab reports, the TA asks the students to think of reasons why this occurred. Fifteen minutes later the students have not thought of any reasons.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 1: Chemistry/ physics version

A group of students has just finished an experiment. They find that their results are the opposite of what they expected. Knowing a question on the week's homework asks students to explain their results, the TA asks the students to think of reasons why this occurred. Fifteen minutes later the students have not thought of any reasons.

Question 2: Biology version

In a lab report, students are asked to develop and defend a physiological explanation for their results. After students hand in their lab reports, their TA reads over their papers and finds that some students did not explain or gave illogical explanations for their results.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Chemistry/physics version

As a homework question, students are asked to identify and explain how an error could have affected their results. After students hand in their homework, their TA reads over their answers and finds that some students did not explain or gave illogical explanations for how the error affected their results.

Nature of Science

Question 1: Generic version

A TA is listening in on a group of students discussing what errors they could have made performing their experiment. The students are confident that they have made a mistake, they are just unsure what they did wrong. The TA asks them why they think they have made a mistake. The students reply that their results must be wrong because they do not match the prediction they made in their hypothesis.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Biology version

A group of students decides to repeat a classical experiment. Their results contradict previous experiments. The group goes to their TA and explains that their experiment did not work because their results do not match past studies.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 2: Physics/chemistry version

Students are asked to repeat a classical experiment. Some groups get results that contradict previous experiments. These groups go to their TA and explain that their experiments did not work because their results do not match past studies.

Group work/ cooperative learning

Question 1: Biology/physics version

On the first day of lab, a TA explains that even though students will be working in groups to complete experiments, every student is individually responsible for knowing the procedures for every part of the experiment. Halfway through the course, the TA notices some students are consistently giving incorrect information for how the experiment was conducted. They asks some of the students why this is the case and they respond that they do not know how to describe parts of experiments that they did not carry out.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

Question 1: Chemistry version

On the first day of lab, a TA explains that even though students will be working in groups to complete some experiments, every student is individually responsible for knowing the procedures for every part of the experiment. Halfway through the course, the TA notices some students are consistently giving incorrect information for how the experiment was conducted for labs that were completed in groups. They asks some of the students why this is the case and they respond that they do not know how to describe parts of experiments that they did not carry out.

Question 2: Generic version

Students are working together to collect data for an experiment. Once the data is collected, one group member completes all necessary calculations and makes a graph of the data. The other student(s) cleans up from the lab. Once the students finish, the student who analyzed the data sends an email with all the calculations and the graph to the rest of the group. The students then leave lab, without discussing their results or how the one students analyzed the data.

As a TA, how often have you experienced something similar to this?

- □ Not applicable
- □ Never
- □ Once a semester
- □ Two times a semester
- □ Three or more times a semester
- □ Almost every week

APPENDIX C

Demographic information for participating TAs (combining BIO113, BIO114)

	# TA Participants
Ethnicity	
White	7
Hispanic/Latino(a)	0
Black	2
Asian	0
Mixed, Undeclared	0
Total	9
Gender	
Female	4
Male	5
Total	9
Teaching Experience PRIOR to WFU	
None	5
Yes; pre-college	2
Yes; college level	2
Total	9
Teaching Experience AT WFU	
None	6
1-2 semesters	2
3-4 semesters	0
> 4 semesters	1
Total	9

Demographic information, including ethnicity, gender and teaching experience is shown for the TAs who taught BIO113 and BIO114 during the Fall 2015 semester. The demographic information is shown for all TAs combined due to the small number of TAs who taught each course.

APPENDIX D

Pre/Post questions about the Nature of Science

Students completed these as part of BASX and end of course evaluation.

*Questions from Views of the Nature of Science Questionnaire Form C (*Lederman et. al., 2002*)*

1. What is an experiment?

2. Does the development of scientific knowledge require experiments?
    a. If yes, explain why. Give an example to defend your position.
    b. If no, explain why. Give an example to defend your position.

*Questions from Views on Science-Technology-Society (*Aikenhead, Ryan, & Fleming, 1989*)*

10112 Defining science is difficult because science is complex and does many things. But MAINLY science is:
Your position, basically:

A. A study of fields such as biology, chemistry, physics
B. A body of knowledge, such as principles, laws and theories, which explain the world around us (matter, energy and life).
C. Exploring the unknown and discovering new things about our world and universe and how they work.
D. Carrying out experiments to solve problems of interest about the world around us
E. Inventing or designing things (for example, artificial hearts, computers, space vehicles)
F. Finding and using knowledge to make this world a better place to live in (for example, curing diseases, solving pollution and improving agriculture)
G. An organization of people (called scientists) who have ideas and techniques for discovering new knowledge
H. I don't understand
I. I don't know enough about this subject to make my choice
J. None of these choices fits my basic viewpoint

90411 Even when scientific investigations are done correctly, the knowledge that scientists discover from those investigations may change in the future.
Your position, basically is:

A. Scientific knowledge changes because new scientists disprove the theories or discoveries of old scientists. Scientists do this by using new techniques or improved instruments, by finding new factors overlooked before, or by detecting errors in the original "correct" investigation.
B. Scientific knowledge changes because the old knowledge is reinterpreted in light of new discoveries. Scientific facts can change.
C. Scientific knowledge APPEARS to change because the interpretation or the application of old facts can change. Correctly done experiments yield unchangeable facts.
D. Scientific knowledge APPEARS to change because new knowledge is added on to old knowledge; the old knowledge doesn't change.

E. I don't understand
F. I don't know enough about this subject to make my choice
G. None of these choices fits my basic viewpoint

90511 Scientific ideas develop from hypotheses to theories, and finally, if they are good enough, to being scientific laws.

Your position, basically:

Hypotheses can lead to theories which can lead to laws:

A. because a hypothesis is tested by experiments, if it proves correct, it becomes a theory. After a theory has been proven true many times by different people and has been around for a long time, it becomes a law.
B. because a hypothesis is tested by experiments, if there is supporting evidence, it's a theory. After a theory has been tested many times and seems to be essentially correct, it's good enough to become a law.
C. because it is a logical way for scientific ideas to develop.
D. Theories can't become laws because they are both different types of ideas. Theories are based on scientific ideas which are less than 100% certain, and so theories can't be proven true. Laws, however, are based on facts only and are 100% sure.
E. Theories can't become laws because they both are different types of ideas. Laws describe things in general. Theories explain these laws. However, with supporting evidence, hypotheses may become theories (explanations) or laws (descriptions).
F. I don't understand
G. I don't know enough about this subject to make my choice
H. None of these choices fits my basic viewpoint

90521 When developing new theories or laws, scientists need to make certain assumptions about nature (for example, matter is made up of atoms). These assumptions must be true in order for science to progress properly.

Your position, basically:

A. Assumptions MUST be true in order for science to progress, because correct assumptions are needed for correct theories and laws. Otherwise, scientists would waste a lot of time and effort using wrong theories and laws.
B. Assumptions MUST be true in order for science to progress; otherwise society would have serious problems, such as inadequate technology and dangerous chemicals.
C. Assumptions MUST be true in order for science to progress, because scientists do research to prove their assumptions true before going on with their work.
D. It depends. Sometimes science needs true assumptions in order to progress. But sometimes history has shown that great discoveries have been made by disproving a theory and learning from its false assumptions.
E. It doesn't matter. Scientists have to make assumptions, true or not, in order to get started on a project. History has shown that great discoveries have been made by disproving a theory and learning from its false assumptions.
F. Scientists do not make assumptions. They research an idea to find out if the idea is true. They don't assume it is true.
G. I don't understand
H. I don't know enough about this subject to make my choice
I. None of these choices fits my basic viewpoint

90611 When scientists investigate, it is said that they follow the scientific method. The scientific method is:

Your position, basically:

A. lab procedures or techniques, often written in a book or journal, and usually by a scientist.
B. recording your results carefully.
C. controlling experimental variables carefully, leaving no room for interpretation.
D. getting facts, theories or hypotheses efficiently.
E. testing and retesting- proving something true or false in a valid way.
F. postulating a theory then creating an experiment to prove it.
G. questioning, hypothesizing, collecting data and concluding.
H. a logical and widely accepted approach to problem-solving.
I. an attitude that guides scientists in their work.
J. Considering what scientists actually do, there really is no such thing as the scientific method.
K. I don't understand
L. I don't know enough about this subject to make my choice
M. None of these choices fits my basic viewpoint

90621 The best scientists are those who follow the steps of the scientific method.

Your position, basically:

A. The scientific method ensures valid, clear, logical and accurate results. Thus, most scientists will follow the steps of the scientific method.
B. The scientific method should work well for most scientists; based on what we learned in school.
C. The scientific method is useful in many instances, but it does not ensure results. Thus, the best scientists will also use originality and creativity.
D. The best scientists are those who may use any method that might get favorable results (including the method of imagination and creativity).
E. Many scientific discoveries were made by accident, and not by sticking to the scientific method.
F. I don't understand
G. I don't know enough about this subject to make my choice
H. None of these choices fits my basic viewpoint

90631 Scientific discoveries occur as a result of a series of investigations, each one building on an earlier one, and each one leading logically to the next one, until the discovery is made.

Your position, basically is that scientific discoveries result from a logical series of investigations:

A. because experiments (for example, the experiments that led to the model of the atom, or discoveries about cancer) are like laying bricks onto a wall.
B. because research begins by checking the results of an earlier experiment to see if it is true. A new experiment will be checked by the people who come afterwards.
C. Usually scientific discoveries result from a logical series of investigations. But science is not completely logical. There is an element of trial and error, hit and miss, in the process.
D. Some scientific discoveries are accidental, or they are the unpredicted product of the actual intention of the scientists. However, more discoveries result from a series of investigations building logically one upon the other.

E. Most scientific discoveries are accidental, or they are the unpredicted product of the actual intention of the scientists. Some discoveries result from a series of investigations building logically one upon the other.
F. Scientific discoveries do not occur as a result of a logical series of investigations:
G. because discoveries often result from the piecing together of previously unrelated bits of information.
H. because discoveries occur as a result of a wide variety of studies which originally had nothing to do with each other, but which turned out to relate to each other in unpredictable ways.
I. I don't understand
J. I don't know enough about this subject to make my choice
K. None of these choices fits my basic viewpoint

90711 Even when making predictions based on accurate knowledge, scientists and engineers can tell us only what probably might happen. They cannot tell what will happen for certain.
Your position, basically is predictions are NEVER certain:
A. because there is always room for error and unforeseen events which will affect a result. No one can predict the future for certain.
B. because accurate knowledge changes as new discoveries are made, and therefore predictions will always change.
C. because a prediction is not a statement of fact. It is an educated guess.
D. because scientists never have all the facts. Some data are always missing.
E. It depends. Predictions are certain, only as long as there is accurate knowledge and enough information.
F. I don't understand
G. I don't know enough about this subject to make my choice
H. None of these choices fits my basic viewpoint

90811 If scientists find that people working with asbestos have twice as much chance of getting lung cancer as the average person, this must mean that asbestos causes lung cancer.
Your position, basically:
A. The facts obviously prove that asbestos causes lung cancer. If asbestos workers have a greater chance of getting lung cancer, then asbestos is the cause.
B. The facts do NOT necessarily mean that asbestos causes lung cancer:
C. because more research is needed to find out whether it is asbestos or some other substance that causes lung cancer.
D. because asbestos might work in combination with other things, or may work indirectly (for example, weakening your resistance to other things which cause you to get lung cancer).
E. because if it did, all asbestos workers would have developed lung cancer.
F. Asbestos cannot be the cause of lung cancer because many people who don't work with asbestos also get lung cancer.
G. I don't understand
H. I don't know enough about this subject to make my choice
I. None of these choices fits my basic viewpoint

91012 For this statement, assume that a gold miner "discovers gold" while an artist "invents" a sculpture. Some people think that scientists discover scientific HYPOTHESES. Others think that scientists invent them. What do you think?

Your position, basically:

Scientists discover a hypothesis:

A. because the idea was there all the time to be uncovered.
B. because it is based on experimental facts.
C. but scientists invent the methods to find they hypothesis.
D. Some scientists may stumble onto a hypothesis by chance, thus discovering it. But other scientists may invent the hypothesis from facts they already know.

Scientists invent a hypothesis:

E. because a hypothesis is an interpretation of experimental facts which scientists have discovered.
F. because inventions (hypotheses) come from the mind- we can create them.
G. I don't understand
H. I don't know enough about this subject to make my choice
I. None of these choices fits my basic viewpoint

Questions about student responsibility

The scale for these questions is based on the work of Zimmerman and Kitsantas (2005).

	1- mainly the teacher	2- definitely more the teacher	3- slightly more the teacher	4- both equally	5- slightly the student	6- definitely more the student	7- mainly the student
Who is more responsible for a student, remembering information from the lab manual?							
Who is more responsible for a student, taking notes in class?							
Who is more responsible for a student, getting clear guidelines for writing lab reports?							
Who is more responsible for a student, arranging meetings with the instructor?							
Who is more responsible for a student, staying awake during lab?							
Who is more responsible for a student, learning the content covered in lab?							
Who is more responsible for a student, writing good lab reports?							
Who is more responsible for a student, participating in class discussions?							
Who is more responsible for a student, receiving timely feedback on their work?							
Who is more responsible for a student, reading assigned materials?							
Who is more responsible for a student, following the directions on assignments?							

APPENDIX E

Frequencies at which TAs in science departments reported instructional challenges

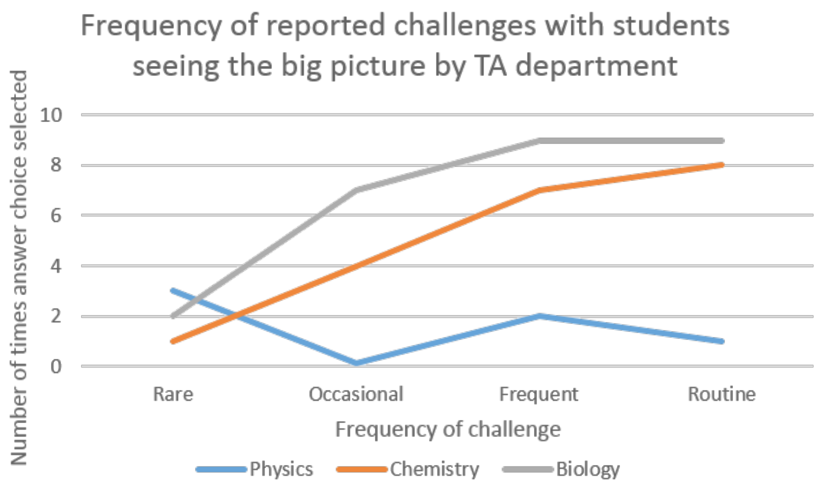

FIGURE 1: How frequently TAs in different science departments reported seeing students struggle with seeing the big picture. On the x-axis, rare represents answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. Y-axis shows number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

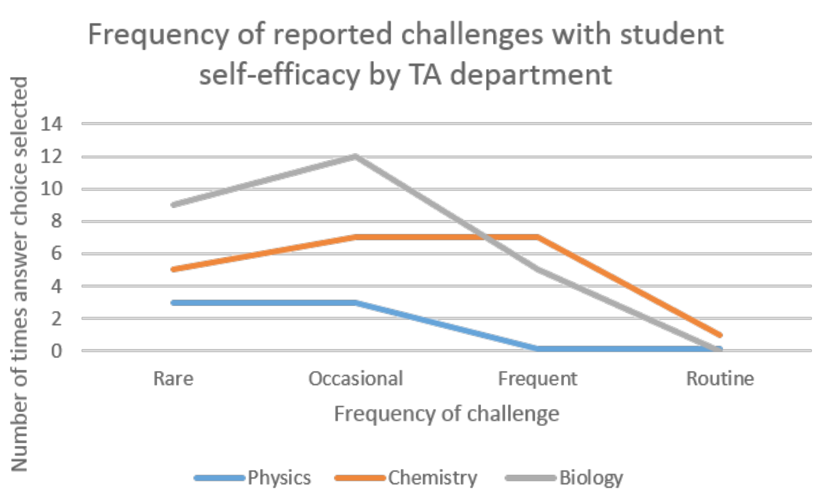

FIGURE 2: How frequently TAs in different science departments reported seeing students struggle with low self-efficacy. On x-axis, rare represents answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. Y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

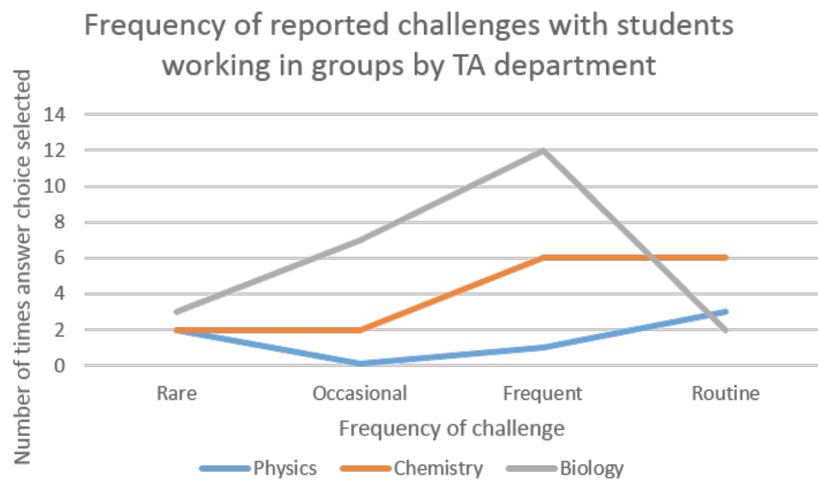

FIGURE 3: How frequently TAs in different science departments reported seeing students struggle with group work. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

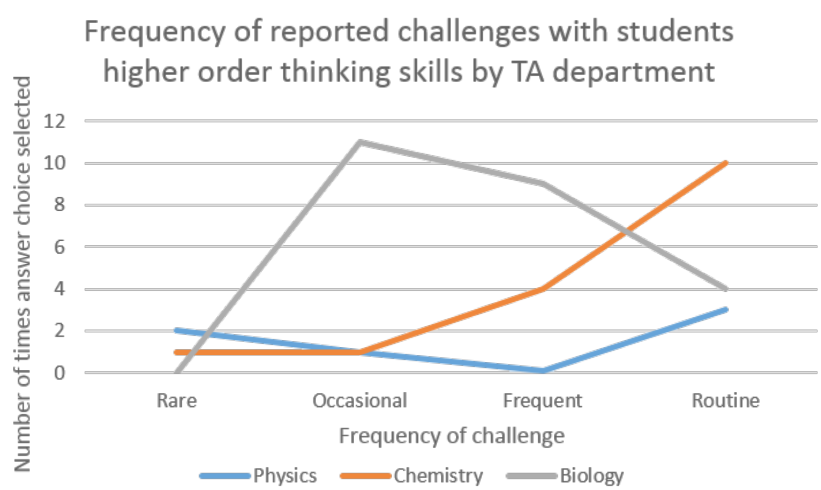

FIGURE 4: How frequently TAs in different science departments reported seeing students struggle with developing higher-order thinking skills. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

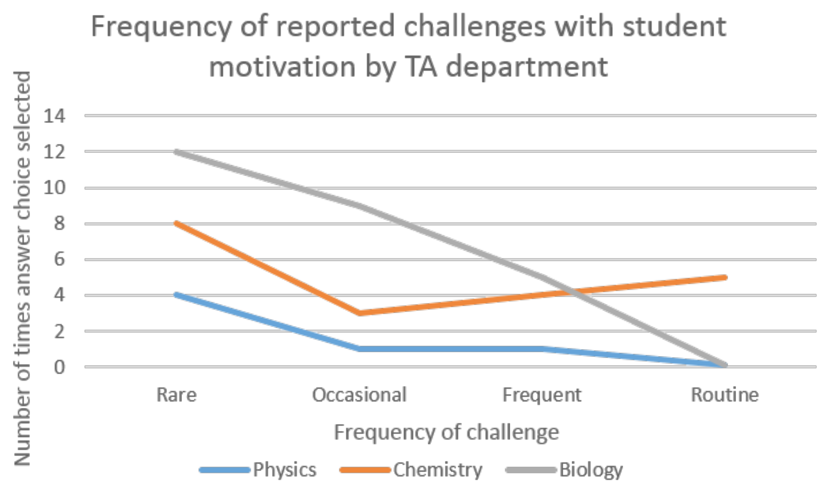

FIGURE 5: How frequently TAs in different science departments reported seeing students struggle with being motivated. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

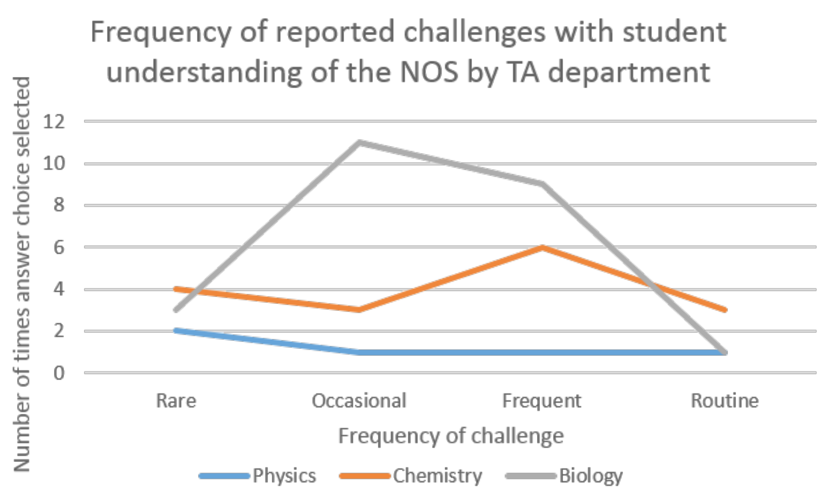

FIGURE 6: How frequently TAs in different science departments reported seeing students struggle with understanding the nature of science. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

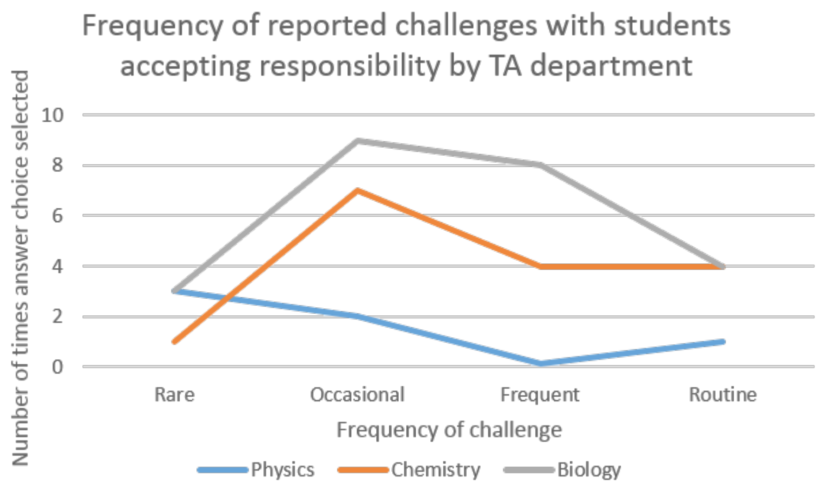

FIGURE 7: How frequently TAs in different science departments reported seeing students struggle with accepting responsibility for their work and performance in the course. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

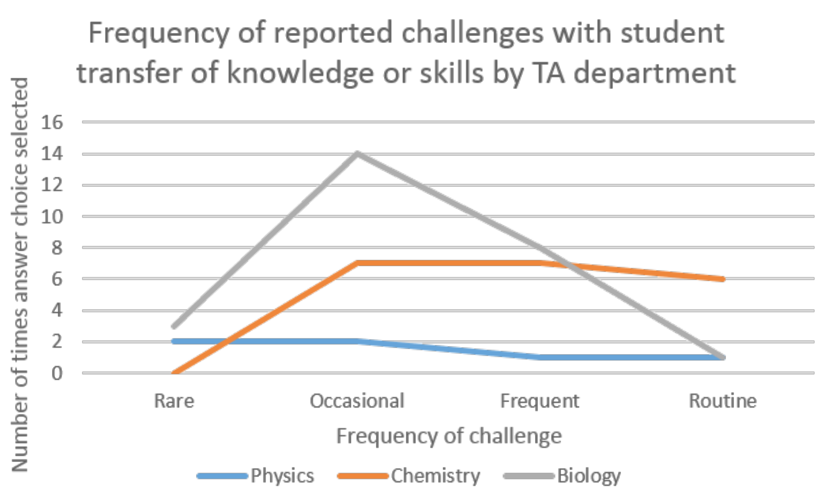

FIGURE 8: How frequently TAs in different science departments reported seeing students struggle with transferring knowledge or skills. On the x-axis, rare represents the answer choices not applicable or never, occasional one or two times a semester, frequent two or more times a semester and routine almost every week. The y-axis shows the number of times each answer choice was selected when the responses to the two questions on each instructional challenge were combined (n=6 for physics, 20 for chemistry, and 24-28 for biology).

APPENDIX F

Learning Log Directions

You will need to complete 2 logs for each of the 4 parts of this class, 1 near the beginning of the course and 1 near the end. These are the log options, and space to write down when they are due.

Class Part	First Log	Due by:	Second Log	Due by:
Class Sessions	LL for Unit II or III		LL for Unit V or VI	
Quizzes	LL for 1 of your first 3 quizzes (Quiz 1, 2, or 3)		LL for 1 of your last 3 quizzes (Quiz 8, 9, or 10)	
Lab Notebooks	LL for Unit II or III		LL for Unit V or VI	
Lab Reports	LL for the first lab report		LL for the final lab report	

How are they graded? When are they due?
You submit your learning logs online. They are graded either Completed or Not Completed. You earn a point for each completed log.

- Class session learning logs are due 48 hours after a class session.
- All other learning logs are due one week after a graded assignment is returned.

What should learning logs contain?
Each assignment has a different set of log questions, which are on the following pages. A log should have:

- Complete answers to every question.
- Each answer should be at least three sentences in length.

Class Session Log

1. What class period are you discussing in this log? Please give a specific date.
2. List the main points you learned from this lab. This can include biology content, procedural information, or details of experimental design. Be specific.
3. List any points from this lab that were unclear to you and at least one topic that you would like to learn more about. Be specific.
4. Only answer 1 of the next 2 options, whichever applies.
    a. If you listed points that were unclear in Question 3, what can you do to get a better understanding of these topics?
    b. If you listed no unclear points in Question 3, what activities, in or out of class, best helped you understand the material (mention at least two). Be specific.

Quiz Log

1. What quiz are you discussing in this log? (Give the number in upper right corner of the quiz)
2. Give an example of one of your most successful responses. Try to explain what things you did that made the answer successful.
3. Only answer 1 of the next 2 options, whichever applies.
    a. If you earned an A or A- on this quiz, explain what things you did to prepare for the quiz?
    b. If you earned a B+ or lower on this quiz, give one example of an error or less successful response. What did you do incorrectly or fail to do when answering this question?
4. Only answer 1 of the next 2 options, whichever applies. Check the box below if you have no prior quiz to compare this one to, or if you earned an A or A- on both this quiz and previous quizzes.
    - □ First quiz
    - □ Earned A/A- on both

    a. *If you earned an A or A- on this quiz*, why do you think you did not perform as well on previous quizzes?
    b. *If you earned a B+ or lower on this quiz*, what could you do differently to increase your learning for the next quiz?

Lab Report or Lab Notebook Log

1. What lab report or lab notebook entry are you discussing in this log?
2. Give an example of something you did well throughout your lab report or lab notebook entry. Explain what things you did to strengthen this part of your paper.
3. List three areas of your lab report or lab notebook entry that need improvement. Be specific.
4. The next time you write a lab report or lab notebook entry, what could you do differently to increase the quality of your assignment? Be specific.

APPENDIX G

Higher-order thinking assignments

Directions for writing a physiological explanation in a lab report

For lab report 1
[The following are a part of a checklist of required criteria for the lab report's discussion]

- Physiological explanation for how injecting a specific compound affected a specific part of muscle contraction. Give evidence supporting this explanation.

- Compare your explanation to the results and interpretations of previously published studies.

For lab report 2
[The following are a part of a checklist of required criteria for the lab report's discussion]

- Physiological explanation for how changing a specific environmental factor affected metabolic rate. Give evidence supporting this explanation.

- Compare your explanation to the results and interpretations of previously published studies.

Quiz questions

Generic directions for all questions:
When asked to make a prediction or give an explanation:
- Make sure you give a plausible prediction or explanation that fits with accepted scientific knowledge.

- Explain your reasoning and provide evidence to support your answer.

From Quiz 1
- Comparative physiology explores how different organisms function. It includes comparing how different organisms obtain energy or comparing at what organs different organisms possess and how these organs work together. Why do you think this field is important? Why do you think we require you to take a course in comparative physiology?

- Scientists spend a lot of time writing. Besides documenting their experimental methods and results, they write formal scientific articles that they submit to scientific journals for publication. Why do you think it is important for scientists to publish their research? Provide at least two reasons.

From Quiz 2

- If you have ever held a frog, you probably noticed their skin is wet and slippery. This is because frogs secrete a mucus that keeps their skin moist when they are out of water. Given what you learned in lab last week about how frogs exchange gases with the environment, what might be the benefit of a frog keeping its skin moist when it is not submersed in water?

- What do you think is the relationship between the relative length and complexity of an organism's digestive tract and the variety (or lack of variety) in an organism's diet? Use examples from the organisms your group dissected last week.

- What are spiracles and where are they located on *Manduca sexta*? What would you expect to happen to gas exchange if one of these spiracles was damaged or blocked? What is your reasoning?

From Quiz 3

In 2009, Rybcxynski et al. investigated whether prothoracicotropic hormone (PTTH) was present in the brains of *Manduca sexta* after pupation. Their paper did not contain an explicitly stated hypothesis, but from their introduction, the following hypothesis is implied: If PTTH has a function other than causing the release of ecdysone in *Manduca sexta*, then PTTH levels will not decline after pupation. Design an experiment to test this hypothesis. (Rybzynski, R., C. A. Snyder, J. Hartmann, L. I. Gilbert (2009). *Manduca sexta* prothoracicotropic hormone: evidence for a role beyond steroidogenesis. *Archives of Insect Biochemistry and Physiology* 70: 217-229. DOI: 10.1002/arch.20295.)

- Imagine you conduct the experiment you described. You find that PTTH levels decline after pupation. What would you conclude about the function of PTTH (make sure you relate your answer to Rybcxynski's hypothesis)?

From Quiz 4

- What is the hypothesis your group developed? Why did you predict adding the hormone at this stage in *Manduca's* life cycle would have this effect?

- At this point in your experiment, do your data support or not support your hypothesis? What can you conclude about the role of the hormone you studied in the life cycle of *Manduca sexta*?

From Quiz 5

Plants grown under green light exhibited approximately 5-fold lower mean transpiration rates than plants grown under white light (Figure 1). Plants in the experimental group had transpiration rates of .000, .001, and .002. Plants in the control group had transpiration rates of .004, .006, and .007.

- Based on what you learned about technical writing last week, how could the following results section be improved? Make sure you evaluate what is present and what is missing.

- What can you conclude from the data above?

From Quiz 6

- Humans consider some insects pests. The tobacco hornworm you worked with in Unit II is considered a pest in its larval stage. Mosquitos are pests as adults. Given what you know about the role of juvenile hormone and ecdysone in insect development, how could you stop tobacco hornworms from being pests? How could you stop mosquitos from being pests?

- As you learned in Unit I, the gills of a crayfish are attached to its legs. What is the functional advantage of having gills in this location?

From Quiz 7

- What do you predict will be the effect of adding acetylcholinesterase inhibitor to the gastrocnemius muscle of *Rana pipiens*?

- What do you predict will be the effect of adding caffeine (which causes calcium ions to be released from the sarcoplasmic reticulum) to the gastrocnemius muscle of *Rana pipiens*?

From Quiz 8

The following is an excerpt from the discussion section of a lab report.

> "The data did not support our hypothesis. We hypothesized that NyQuil would increase heart rate, but our data show the opposite. This could be because we based our predictions on the function of doxylamine succinate, which is only one ingredient in NyQuil. NyQuil contains many other ingredients, which canceled out the effect of doxylamine in our experiment. Ethanol was likely one of these ingredients. NyQuil contains about 10% ethanol. Kaas et al. (2009) found that a 5% ethanol solution decreased heart rate in *Daphnia magna* after 2 minutes and that heart rate remained decreased at 10 minutes. They did not measure heart rate at 15 minutes, but if a 5% ethanol solution caused a decrease in heart rate at 10 minutes, it seems reasonable that a 10% ethanol solution would cause a significant decrease at 15 minutes."

- What explanation did the author give for why their data contradicted their hypothesis?

- What could the author do to test this explanation?

From Quiz 9

- What do you predict would happen to the metabolic rate of an aquatic endotherm if you increased the temperature of the surrounding water?

- An aquatic ectotherm is placed in a dark tank for 12 hours and then its metabolic rate is measured. Do you predict its metabolic rate would be higher or lower than an aquatic ectotherm exposed to light for 12 hours? What is your reasoning?

APPENDIX H

Higher-order thinking scoring tools

Generic scoring tool from (McNeill & Krajcik, 2011):

	Claim: A statement or conclusion that answers the original question/ problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim.	Does not provide evidence.	Does not provide reasoning.
1	Makes an accurate and complete claim.	Provides vague or inappropriate evidence that does not support the claim.	Provided reasoning is inappropriate and does not link evidence to claim.
2	N/A	Provides evidence, but the evidence is insufficient. Provides at least one but not all of the following pieces of evidence:	Provides reasoning, but the reasoning is insufficient. Provides at least one but not all of the following reasoning components:
3	N/A	Provides all the following pieces of evidence:	Provides all the reasoning components:

Scoring tool for lab reports

	Claim: A statement or conclusion that answers the original question/ problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim. e.g. Summarizes results.	Does not provide evidence.	Does not provide reasoning. E.g. only summarizes background information and a previous study.
1	Does not explicitly make a claim, but describes physiological background in addition to results.	Provides vague or inappropriate evidence that does not support the claim.	Provided reasoning is inappropriate and does not link evidence to claim.
2	Makes an incomplete claim (may not be explicitly stated) related to the original question/ problem (explanation of the results). E.g. Explains why the results were not significant but does link this to what was learned in class (physiology).	Provides evidence, but the evidence is insufficient. Provides at least one but not all of the following pieces of evidence: - Data, experimental methods, errors or limitations - Previous studies (must explain the methods and results of the study not just use it as a source of information).	Provides reasoning, but the reasoning is insufficient. Provides at least one but not all of the following reasoning components: - Gives plausible connection between data, experimental methods, errors or limitations and claim - Gives plausible connection between previous studies and claim NOTE: At least one of these connections should be based on physiological concepts.
3	Explicitly states a complete claim (explaining results) that considers physiological concepts.	Provides all the evidence necessary to support the claim, as listed above.	Provides all the reasoning necessary to support the claim, as listed above.

Scoring tool for question 1 on quiz 2

	Claim: A statement or conclusion that answers the original question/ problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim. E.g. describes respiratory system	Does not provide evidence.	Does not provide reasoning.
1	Makes a vague claim or claim that is not related to the original question/problem (benefit of moist skin relating to gas exchange). E.g. Breathing through the skin is the most effective way to breath.	Provides only or almost all vague, incorrect, or inappropriate evidence or evidence that does not support the claim.	Provided reasoning is only or almost all inappropriate, incorrect or vague, or does not link evidence to claim.
2	Makes an incomplete claim. E.g. The benefit is that frogs can breathe through their skin while on land (does not take lungs into account). Note: Answers that give two claims, at least one of which is accurate, go in this category.	Provides evidence, but the evidence is insufficient. Provides one of the following pieces of evidence: - A frog's skin is very thin and immediately under the skin are blood vessels (as observed in lab). - Frogs have lungs they use to breathe when on land (as observed in lab). Or other additional or alternative evidence as needed to support the claim. E.g. Evidence that skin needs to stay hydrated on land to conduct gas exchange in water (if this is claimed to be the benefit).	Provides reasoning, but the reasoning is insufficient. Provides at least one but not all of the following reasoning components: - Respiratory surfaces must be moist for gas exchange to occur. - Frogs are amphibians and must have a mechanism for breathing while underwater and on land. - When frogs exchange gases through their skin this is a passive process that does not require energy input. Breathing through the lungs requires energy. OR other reasoning as needed to support the claim.
3	Provides an explicitly stated complete claim. E.g. Frogs secrete a mucus to keep their skin moist so they carry out gas exchange through their skin, even while on land. Gives a reason for why gas exchange may occur through the skin in addition to or in lieu of the lungs.	Provides all the evidence necessary to support the claim, as listed above.	Provides all the reasoning necessary to support the claim, as listed above.

Scoring tool for question 2 on quiz 2

	Claim: A statement or conclusion that answers the original question/problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim.	Does not provide evidence.	Does not provide reasoning.
1	Makes a vague claim or claim that is not related to the original question/ problem (relationship between variety of diet and length and complexity of digestive tract). E.g. There is a direct relationship.	Provides only or almost all vague, incorrect, or inappropriate evidence or evidence that does not support the claim. E.g. the digestive tracts are more or less complex.	Provided reasoning is only or almost all inappropriate, incorrect or vague, or does not link evidence to claim.
2	Makes an incomplete claim. E.g. The relative length of an organism's digestive tract increases as the variety in the organism's diet increases.	Provides evidence, but the evidence is insufficient. Provides two of the following pieces of evidence: - The frog had a digestive tract 1.5-2x its body length. - The frog also had the most accessory organs. - The caterpillar and/or crayfish had digestive tracts that were about the same length as their body. - The crayfish and/or frog had few or no accessory organs.	Provides reasoning, but the reasoning is insufficient. Provides the following reasoning components: - Frogs have the most varied diet, ranging from other smaller frogs to insects. - The caterpillars we studied eat only tobacco (okay to also say tomato) leaves and/or the crayfish eat decaying matter. - In place of two above points, may say diet of frogs are more varied and others are less varied. Does not give an accurate scientific principal that explains why a more varied diet is linked to a longer and more complex digestive tract.
3	Makes a complete claim. E.g. The relative length and complexity of an organism's digestive tract increases as the variety in the organism's diet increases.	Provides all the following pieces of evidence: - The frog had a digestive tract 1.5-2x its body length. It also had the most accessory organs. - The caterpillar and crayfish had digestive tracts that were about the same length as their body and had few accessory organs.	Provides all the reasoning components: - Frogs have the most varied diet, ranging from other smaller frogs to insects. - The caterpillars we studied eat only tobacco (okay to also say tomato) leaves and the crayfish eat decaying matter. - An accurate scientific principal that explains why a more varied diet is linked to a longer and more complex digestive tract.

Scoring tool for question 3 on quiz 2

	Claim: A statement or conclusion that answers the original question/problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim.	Does not provide evidence.	Does not provide reasoning.
1	Makes a vague claim or claim that is not related to the original question/problem. e.g. Gas exchange would be impaired.	Provides only or almost all vague, incorrect or inappropriate evidence that does not support the claim.	Provided reasoning is only or almost all inappropriate, incorrect or vague, or does not link evidence to claim.
2	Makes an incomplete claim (does not consider what would happen to gas exchange overall). E.g. The *Manduca* will be able to survive. OR Gas exchange would stop at the damaged spiracle and continue at the others.	Provides evidence, but the evidence is insufficient. Provides one of the following pieces of evidence: - *Manduca* have many spiracles. - The tracheal tubes that come from each spiracle fan out and overlap, carrying oxygen to more than one body segment. Or other additional or alternative evidence as needed to support the claim. E.g. Evidence that spiracles can bring in more oxygen and compensate for a damaged spiracle (if this is claimed).	Provides reasoning, but the reasoning is insufficient. Provides one of the following reasoning components: - *Manduca* have many spiracles where gas exchange can occur. - O_2 is transported to the rest of the body and CO_2 removed by tracheal tubes. - These tubes overlap, so O_2 from different spiracles can reach the same body segment. OR other reasoning as needed to support the claim.
3	Makes a complete claim that considers what would happen to gas exchange overall (decrease, increase, stay the same).	Provides all the evidence necessary to support the claim, as listed above.	Provides all the reasoning necessary to support the claim, as listed above.

Scoring tool for question 1 on quiz 6

	Claim: A statement or conclusion that answers the original question/problem.	**Evidence:** Scientific data or observations that supports the claim. The data needs to be appropriate and sufficient to support the claim.	**Reasoning:** A justification that connects the evidence to the claim. It shows why the data count as evidence by using appropriate and sufficient scientific principles.
0	Does not make a claim.	Does not provide evidence.	Does not provide reasoning.
1	Makes a vague claim or claim that is not related to the original question/ problem (how to stop tobacco hornworms and mosquitos from being pests).	Provides only or almost all vague, incorrect, or inappropriate evidence or evidence that does not support the claim.	Provided reasoning is only or almost all inappropriate, incorrect or vague, or does not link evidence to claim.
2	Makes an incomplete claim. E.g. You should add ecydsone to *Manduca* (does not say when the ecdysone should be added).	Provides evidence, but the evidence is insufficient. Provides one of the following pieces of evidence: - A spike in ecydsone in the presence of JH causes a molt. - A spike in ecydsone in the absence of JH causes pupation.	Provides reasoning, but the reasoning is insufficient. Provides some of the following reasoning components: - Removing JH when ecydsone spikes will cause pupation. - Adding JH when it is normally not around can prevent pupation.
3	Makes a complete claim that considers the life stage of each organism. Note: It is OK if the question only discusses how to keep one organism from being a pest.	Provides all the evidence necessary to support the claim, as listed above.	Provides all reasoning necessary to support the claim, including the two components listed above and the following: - The tobacco hornworm needs to become an adult to no longer be a pest and the mosquito needs to stay a larva to not become a pest.

APPENDIX I

Frequency of specific answer choices on VOSTS questions

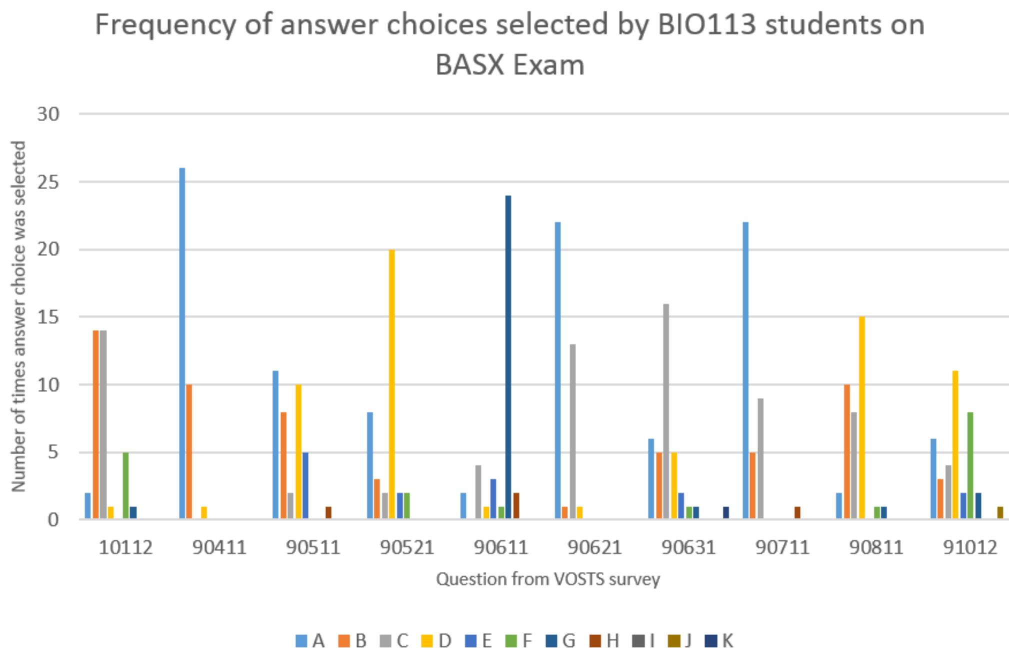

FIGURE 1: The frequency of each answer choice selected by BIO113 students on the BASX exam administered on the first day of the BIO113 lab. The x-axis gives the unique identifying number for each question from the VOSTS survey and the y-axis shows the number of times each answer choice was selected (n=37 students). Answer choices L and M were left out of the figure because no students selected these answer choices.

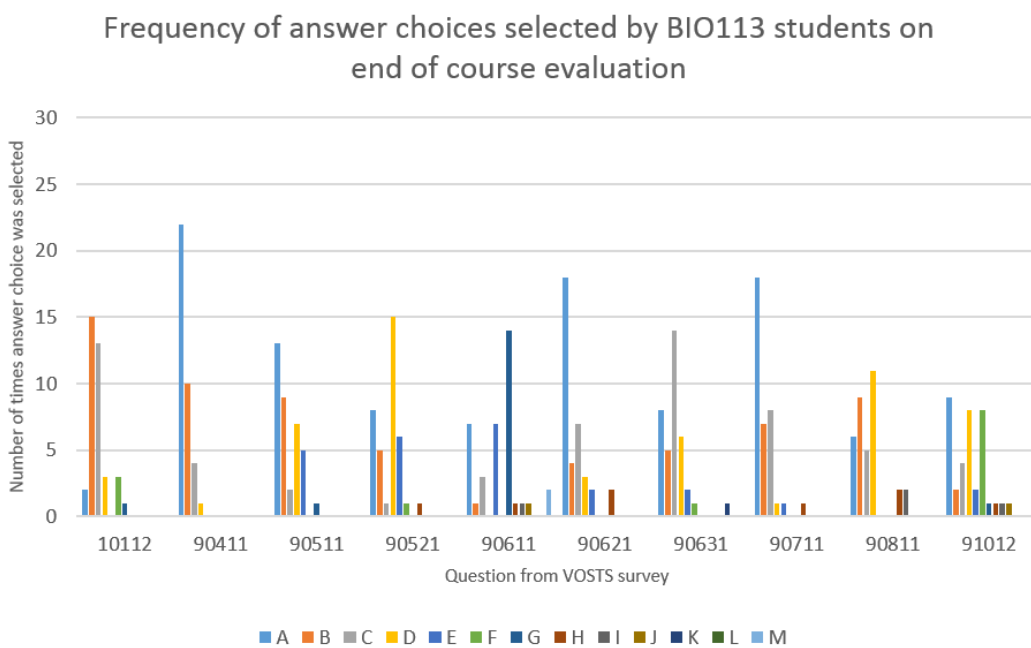

FIGURE 2: The frequency of each answer choice selected by BIO113 students on the end of course evaluation administered on the last day of the BIO113 lab. The x-axis gives the unique identifying number for each question from the VOSTS survey and the y-axis shows the number of times each answer choice was selected (n=37 students).

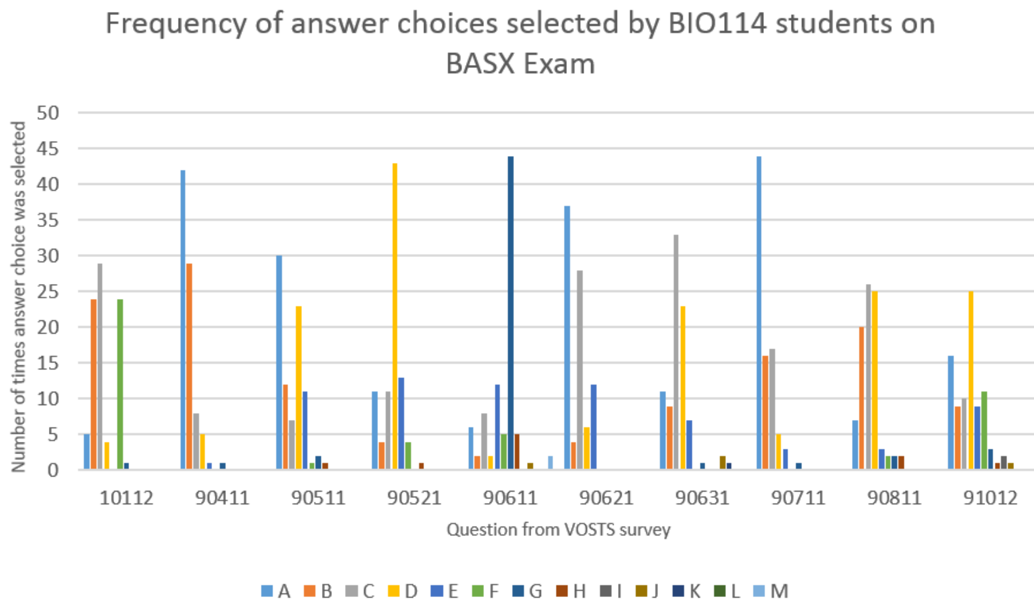

FIGURE 3: The frequency of each answer choice selected by BIO114 students on the BASX exam administered on the first day of the BIO114 lab. The x-axis gives the unique identifying number for each question from the VOSTS survey and the y-axis shows the number of times each answer choice was selected (n=87 students).

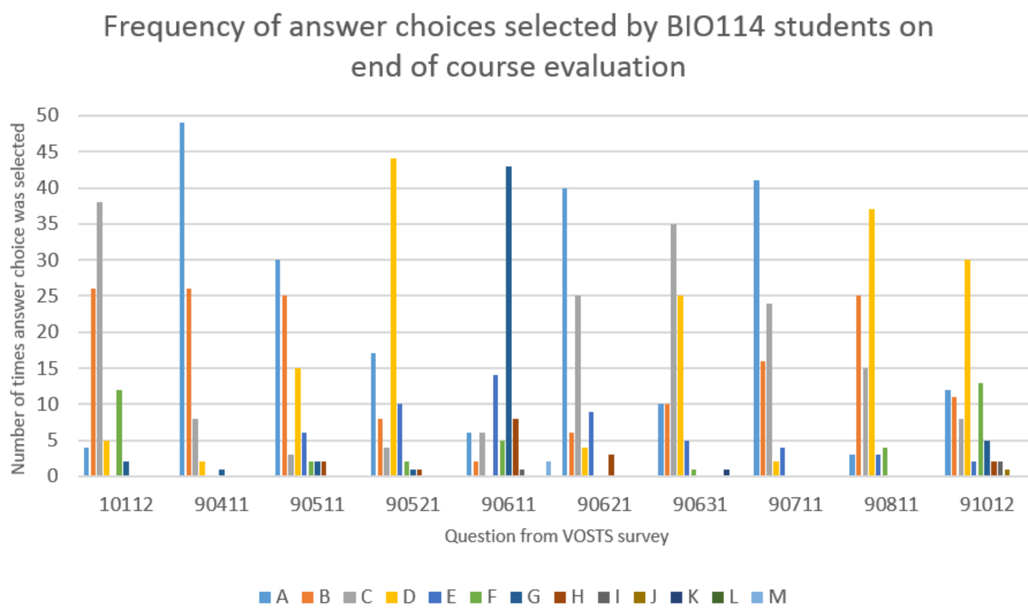

FIGURE 4: The frequency of each answer choice selected by BIO114 students on the end of course evaluation administered on the last day of the BIO114 lab. The x-axis gives the unique identifying number for each question from the VOSTS survey and the y-axis shows the number of times each answer choice was selected (n=87 students).

CURRICULUM VITAE

Jessica Blackburn Martin

Dept. of Biology, Wake Forest University
1834 Wake Forest Road
Winston–Salem, NC 27109
Phone: 828-817-6932
Email: blacjl0@wfu.edu
LinkedIn: https://www.linkedin.com/in/jessica-martin-3678b163?trk=hp-identity-name

EDUCATION:

2014-2016	Wake Forest University, Winston-Salem, NC. M.S. in Biology (expected)
2010-2013	Wake Forest University, Winston-Salem, NC. B.S. in Biology, Magna Cum Laude

PROFESSIONAL POSITIONS:

2014	Research Assistant, Dept. Biology, Wake Forest University, Winston-Salem, NC.
2014	Math and Science Tutor, Sloan Academics, Winston-Salem, NC.

TEACHING EXPERIENCE:

2014-2016	Teaching Assistant, BIO114L “Comparative Physiology Lab”, Wake Forest University, Winston-Salem, NC.

PUBLICATIONS:

Chrispeels, HE, Klosterman, ML, **Martin, JB,** Lundy, SR, Watkins, JM, Gibson, GL, Muday, GK (2014) Undergraduates achieve learning gains in plant genetics through peer teaching of secondary students. *CBE Life Sci Educ.* 13(4): 641-52.

POSTER PRESENTATIONS:

Martin, JB, Johnson, AD. 2015. Undergraduate Biology Education Research: Gordon Research Conference, Lewiston, ME. “Addressing instructional challenges identified by TAs to improve outcomes for undergraduate students in introductory biology labs.”

PROFESSIONAL MEMBERSHIP:

2015-present	National Science Teachers Association
2015-present	North Carolina Science Teachers Association

OTHER POSITIONS AND SERVICE:

2014-2016	Assistant Director, Winston-Salem Regional Science Olympiad Tournament, Winston-Salem, NC.
2013	Media Coordinator, Winston-Salem Regional Science Olympiad Tournament, Winston-Salem, NC.
2013	Communications and External Relations Intern, News Team, Wake Forest University, Winston-Salem, NC.
2013	Volunteer Teacher and Assistant, “Teaching with Tomatoes” Project, Wake Forest University, Winston-Salem, NC.

2013	Undergraduate Research Fellow, Dept. of Biology and English, Wake Forest University, Winston-Salem, NC.
2010-2012	Co-author and Lead Student Reviewer, *The Adapa Project,* Wake Forest University, Winston-Salem, NC.
2011	Teaching Assistant, First Year Seminar “Life’s Molecular Machines”, Wake Forest University, Winston-Salem, NC.

SUPPLEMENTAL COURSEWORK:

2014	Statement of Accomplishment with Distinction in the course “An Introduction to Evidence-based Undergraduate STEM Teaching”, Vanderbilt University, University of Wisconsin–Madison, Michigan State University & Boston University, Coursera Inc.
2012	Statement of Accomplishment with Distinction in the course “Writing in the Sciences”, Stanford University, Coursera Inc.