Michigan Technological University

Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2017

ASSESSING ASSESSMENTS: A COMPARATIVE STUDY OF THREE ASSESSMENTS OF INQUIRY SKILLS IN TENTH GRADE HIGH SCHOOL BIOLOGY STUDENTS

Heather K. Murphy
Michigan Technological University, hkmurphy@mtu.edu

Recommended Citation

Murphy, Heather K., "ASSESSING ASSESSMENTS: A COMPARATIVE STUDY OF THREE ASSESSMENTS OF INQUIRY SKILLS IN TENTH GRADE HIGH SCHOOL BIOLOGY STUDENTS", Open Access Master's Report, Michigan Technological University, 2017.
https://digitalcommons.mtu.edu/etdr/449

Follow this and additional works at: https://digitalcommons.mtu.edu/etdr

Part of the Educational Assessment, Evaluation, and Research Commons

ASSESSING ASSESSMENTS: A COMPARATIVE STUDY OF THREE ASSESSMENTS OF INQUIRY SKILLS IN TENTH GRADE HIGH SCHOOL BIOLOGY STUDENTS

By

Heather K. Murphy

A REPORT

Submitted in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

In Applied Science Education

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

This report has been approved in partial fulfillment of the requirements for the Degree of MASTERS OF SCIENCE in Applied Science Education.

Department of Cognitive and Learning Sciences

Report Co-Advisor: *Shari Stockero*

Report Co-Advisor: *Joshua Ellis*

Committee Member: *Amy Lark*

Department Chair: *Susan Amato-Henderson*

Table of Contents

List of Figures

vi

List of Tables

Abstract

Inquiry is an educational technique where students develop scientific knowledge and understanding through processes similar to those used by scientist. One of the obstacles to implementing the technique is the need for teachers to self-evaluate their teaching to determine the effectiveness of inquiry-based teaching techniques with respect to developing higher-order thinking (HOT) skills in students. Feedback from an assessment permits teachers to adjust their curriculum to incorporate the most effective inquiry-based lessons.

This report considers three short assessments of inquiry that were modified for use in the high school biology classroom. First these three assessments were compared against one another for equality and applicability. Then, each of these three assessments were compared to a longer performance-based assessment for the application to evaluate HOT skills. The purpose of comparing these three assessments to the performance-based assessment was to determine the degree to which each assessment measures HOT skills. Two assessment tools were determined to conform to the requirements of being equally accessible to all students, applicable to being realistic and manageable to use in the classroom and able to evaluate HOT skills.

Chapter 1 Introduction

For me, science education is more than the memorization of vocabulary; it is developing the ability to use higher-order thinking (HOT) skills to solve problems and the ability to apply content-specific knowledge to think logically about issues that the next generation will encounter. In other words, the goal of science education is to help develop scientifically literate individuals who are prepared for the global problems related to science, health, and technology (Erdogan, Campbell, & Abd-Hamid, 2011; Harlen, 2000; National Research Council [NRC] (2012); Next Generation Science Standards (NGSS) Lead States, 2013; Roseman et al., 2016; Rutherford & Ahlgren, 1990). To accomplish this goal, there are many instructional techniques that can be used, and one of them is inquiry-based learning. I will be focusing on inquiry as my innovative instructional technique to develop HOT skills (such as problem solving) within my students.

Inquiry became part of the educational vocabulary sometime in the early twentieth century (Barrow, 2006) and has been a significant theme in my continuing education at Michigan Technological University. Inquiry is an educational technique where students develop scientific knowledge and understanding through processes similar to those used by scientists (NRC, 1996; NGSS Lead States, 2013). Students engage in inquiry by participating in learning activities that can cover a variety of subject matter depending on the focus of the class. Inquiry is valuable to my instruction as it can initiate curiosity when students engage in the process. Observing students engaged in

inquiry also allows me to gain insight into the students' thought process so that misconceptions can be confronted and HOT skills can be developed.

As noted by several researchers (e.g., Harrison, 2014; Hofstein, Novon, Kipnis & Mamlok-Naaman, 2005), one of the obstacles to the implementation of inquiry in the classroom is the teacher's need for assessment tools that evaluate HOT skills. Feedback from an assessment tool permits teachers' to adjust their curriculum to incorporate the most effective inquiry-based teaching techniques.

This report will explore a comparison of three different short assessments, modified from past research. First these three assessments will be compared against one another for equality and applicability. Then, each of these three assessments will be compared to a longer performance-based assessment for the application to evaluate HOT skills. The performance-based approach involved assessing the students over the course of an entire semester-long inquiry process. The three shorter assessments were administered during one 55 minute class period. The purpose of comparing these three assessments to the performance-based assessment is to determine the degree to which each assessment measures HOT skills. The ideal short assessment will be capable of predicting similar results to that of the longer performance-based assessment. This is part of a continuing effort to find a tool that will enable teachers to self-evaluate their teaching techniques to evaluate HOT skills in a timely and valid way.

A quality assessment tool should also be equally accessible to all students, applicable to the classroom where it will be used and able to evaluate the content or skills

of interest (Webb, 1997). An assessment tool is equally accessible if it is similarly available to low achieving students as it is to higher achieving students. Structurally, the assessment needs to be applicable for the tenth grade high school biology classroom. A classroom assessment is system applicable if it requires a reasonable amount of materials, is amenable to implementing within one class period of 55 minutes, and is able to be easily evaluated with little to no ambiguity and within a reasonable time. It is important that the assessment is applicable so that it meets the budgetary and time restrictions inherently present in my tenth grade biology classroom. It is important that the assessment evaluates appropriate content or skills, in the case of this study, HOT skills. If the assessment only evaluates lower-order (LOT) memorization it will not give an accurate depiction of the lesson's ability to develop HOT skills.

Research Question

The intent of this report is to evaluate the assessments to determine whether they meet the criteria above and in order to identify which can be used as a self-evaluation tool to evaluate the effectiveness of inquiry-based teaching techniques with respect to developing HOT skills in students. This report will compare students' scores on three assessments tools, developed by other researchers and modified for a tenth grade biology class, to student scores on a more time intensive performance-based assessment. The specific question that the study addresses is: Of the three assessments (that are called Assessment 1, Assessment 2, and Assessment 3) which demonstrate the following characteristics?

ASSESSING ASSESSMENTS

1. Ability to evaluate higher-order thinking skills

2. Equality and fairness

3. System applicability (realistic and manageable)

Chapter 2 Literature Review

Inquiry in the Science Classroom

Scientific knowledge is accumulated through the scientific processes of synthesis, application, and evaluation (Saido, Siraj, Nordin, Bakar, & Al Amedy, 2015). Inquiry is an instructional technique where students study the natural world using these scientific processes. Through inquiry, students accumulate scientific knowledge and understanding of scientific ideas by experiencing how scientists study the natural world (Hasse, Joachim, Bögeholz, & Hammann, 2014). According to Barrow (2006), as early as 1910, connections were made between developing HOT skills and incorporating inquiry into the classroom. According to the NRC (2000), the five essential features of classroom inquiry are:

> (a) learners are engaged by scientifically oriented questions, (b) learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions, (c) learners formulate explanations from evidence to address scientifically oriented questions, (d) learners evaluate their explanations in light of alternative explanations, (e) learners communicate and justify their proposed explanations. (p. 25)

When these essential features are implemented the inquiry technique puts the student at the heart of the learning and scientific processes (NRC, 2000).

The vast benefits of incorporating inquiry teaching techniques into the science classroom have been documented by many (e.g., Harlen, 2000; Erdogan, Campbell, &

Abd-Hamid, 2011; Quellmalz, Timms, Silberglitt, & Buckley, 2012; Puncochar & Klett, 2013). Zion and Medelovici (2012) emphasize in their model for implementing inquiry— based on an Israeli high school biology inquiry program that had been in place for twelve years—that teaching through inquiry develops scientific literacy, critical thinking, logic, and creative thinking among students. The *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning* (NRC, 2000) as well as by *Project 2061* (Rutherford & Ahlgren, 1990) both all emphasize the importance of science education through inquiry because it helps students expand their skills and abilities, develop long term understanding of science, and become more science literate.

The NRC (2013) publication *Next Generation Science Standards: For States, by States* defines inquiry-based instruction as an approach of teaching science that allows students to be immersed in the practice of science and not "merely" learning about science secondhand (p. xv). The document expresses that by implementing the Next Generation Science Standards (NGSS), strong science-based skills related to content areas, critical thinking, and inquiry-based problem solving practices will be developed. Similarly the document *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*, also published by the NRC (2012), refers to students developing the ability to practice scientific inquiry. The term inquiry can be interpreted in multiple ways; for the intent of this report, however, inquiry is defined as an educational technique where students develop scientific knowledge and understanding through processes similar to those used by scientists.

ASSESSING ASSESSMENTS

According to the NRC, as recorded in Barrow's (2006) *A Brief History of Inquiry: From Dewey to Standards*, there are six fundamental abilities of inquiry. Table 1 below contains the definition of each of these six fundamental abilities of inquiry. The second column lists the NGSS that is aligned with each of the fundamental abilities (NGSS Lead States, 2013). The third column contains an example of the NGSS for each of the abilities as provided by the NGSS Lead States (2013).

Table 1

The Six Fundamental Abilities of Inquiry (Barrow, 2006)

	National Research Council (NRC)	Next Generation Science Standards (NGSS)	Example from Next Generation Science Standard (NGSS)
1	Identify questions and concepts that guide investigations (student formulate and testable hypothesis and an appropriate design to be used);	Scientific Inquiry is characterized by a common set of values that include logical thinking, precision, open-mindedness, objectivity, skepticism, reliability of results, and honest and ethnical reporting of findings (HS-LS1-3)	Example: HS-LS3-1 Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristics traits passed from parents to offspring
2	Design and conduct scientific investigations (using major concepts, proper equipment, safety precautions, use of technologies, etc., where students must use evidence, apply logic, and construct an argument for their proposed explanations);	Plan and conduct an investigation individually and collaboratively to produce data to serve as the basis for evidence, and in the design; decide on types, how much, and accuracy of data needed to produce reliable measurements and consider limitations on the precision of data and refine the design accordingly. (HS-LS1-3)	Example: HS-LS1-3 Plan and conduct and investigation to provide evidence that feedback mechanisms maintain homeostasis.

ASSESSING ASSESSMENTS

Table 1 (cont.)

	National Research Council (NRC)	Next Generation Science Standards (NGSS)	Example from Next Generation Science Standard (NGSS)
3	Use appropriate technologies and mathematics to improve investigations and communications;	Use mathematical representations of phenomena or design solutions to support claims (HS-LS2-4)	Example: HS-LS2-4 Use mathematical representations to support claims for the cycling of matter and flow of energy among organisms in an ecosystem.
4	Formulate and revise scientific explanations and models using logic and evidence (the students' inquiry should result in an explanation or a model);	Construct an explanation based on valid and reliable evidence obtained from a variety of sources (including student's own investigation, models, theories, simulations, peer review) and the assumption that theories and laws that describe the natural world operate today as they did in the past and will continue to do so in the future. (HS-LS1-1)	Example: HS-LS1-1 Construct and explanation based on evidence for how the structure of DNA determines the structure of proteins which carry out the essential functions of life through systems of specialized cells.
5	recognize and analyze alternative explanations and models (reviewing current scientific understanding and evidence to determine which explanations of the model is best);	Design, evaluated, and refine a solution to a complex real-world problem, based on scientific knowledge, student-generated sources of evidence, prioritized criteria, and tradeoff considerations. (HS-LS2-7).	Example: HS-LS2-7 Design, evaluate, and refine a solution for reducing the impacts of human activities on the environment and biodiversity.
6	Communicate and defend a scientific argument (students should refine their skills by presenting written and oral presentations that involve responding appropriately to critical comments from peers).	Communicate scientific information (e.g., about phenomena and/or process of development and the design and performance of a proposed process or system) in multiple formats (including orally, graphically, textually, and mathematically) (HS-LS4-1)	Example: Communicate scientific information that common ancestry and biological evolution are supported by multiple lines of empirical evidence.

The NRC's six fundamental abilities of inquiry, which as shown in Table 1 align with the NGSS, are all strong science-based skills related to content, critical thinking and inquiry practices that can also be defined as HOT skills. This is the connection between inquiry teaching techniques and current standards for education such as the NGSS.

Assessments and Inquiry-Based Science Education

Researchers (e.g., Rutherford & Ahlgren, 1990; Zion & Medelovici, 2012), along with the NRC (2000, 2012), have established the importance of inquiry for developing knowledge, understanding and scientific literacy skills. One obstacle in the implementation of inquiry into the classroom is the teachers' need for assessment tools that will allow them to evaluate the effectiveness of inquiry-based instruction (Webb, 1997; Puncochar, & Klett 2013). As Harrison (2014) indicated, assessment is still one of the most problematic areas limiting the development of inquiry-based instruction. Hofstein et al. (2005), describe the problem of assessing students' achievement in the inquiry learning environment as a "crucial" problem regarding the implementation of inquiry-based laboratory experiments (p. 803).

Evaluation of Assessments

Assessments for teachers to self-evaluate are critical to inquiry-based instruction because they allow the teacher to sharpen and define the design of the learning experience (NRC, 2000). Webb (1997) defines major criteria for evaluating assessments. Two of these criteria pertain to a summative assessment for self-evaluation: equality and fairness, and system applicability.

ASSESSING ASSESSMENTS

Webb (1997) defines equality and fairness as giving "every student a reasonable opportunity to demonstrate attainment of what is expected." (p. 25). Specifically, the level of prior knowledge, culture (ethnic, racial, and gender), social background, and experiences should not contribute to the student's ability to perform on an assessment. Moss (1992) summarizes the definition of fairness as the equitable access to resources such as calculators, the lack of gaps between individuals derived from differences in exposure, and motivation. Fairness also refers to the absence of offensive materials or requirements of unevenly distributed prior knowledge creating a bias (Moss, 1992). However, Webb (1997) explains that "it may be difficult to gauge alignment between expectations and assessments on the criteria of fairness and equality until both have been in place for some time" (p. 27).

An applicable assessment, according to Webb (1997), should be "realistic and manageable" and "can be used by teachers and administrators in a day-to-day setting" (p. 30). An applicable assessment would be one that does not require more than one class period (approximately 55 minutes) to administer. The assessment must also be readily interpreted or gradable. The ideal assessment for this study would have to be realistic and manageable for an individual teacher to implement, and therefore the assessments in this study will be evaluated for system applicability.

The third assessment criteria is content, which in this study will be HOT skills as defined by Bloom's taxonomy (Bloom, Englehart, Furst, & Hill, 1956). Bloom's taxonomy was developed in the first half of the twentieth century and is still considered

to be a significant and appropriate tool for distinguishing learning objectives (Bacon, 2003; Clark, 2010). The Taxonomy of Education Objectives (Bloom, 1956) is a framework for classifying intellectual skills and abilities into six categories listed here in increasing order of complexity: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. The first two levels, Knowledge and Comprehension, are considered LOT skills and the upper four levels, Application, Analysis, Synthesis, and Evaluation, are considered HOT skills.

Common Types of Assessments in the Classroom

Assessments can be broadly classified into two types: traditional paper-and-pencil methods and performance-based assessment. The traditional paper-and-pencil methods include but are not limited to multiple-choice, true/false, free-response, essay, fill in the blank, and multiple-choice with justification. The performance-based assessment formats include but are not limited to journaling, portfolios, scoring rubrics, skill tasks, extended investigation, graphic organizers, concept maps, oral presentations, interviews and conferences. The three assessments developed in this report are short paper-and-pencil format assessments. They are to be compared against a longer performance-based assessment to determine if similar results can be obtained.

The Debate over Multiple-choice Questions

Of the traditional paper-and-pencil methods, multiple-choice formats seem to be employed frequently but are quite controversial (Bridgeman & Rock, 1993; Bridgeman & Lewis, 1994; Lukhele, Thissen & Wainer, 1994; Kohn, 2000; Bacon, 2003). Advantages

11

of multiple-choice format testes include the economic advantages, but it is debated as to whether multiple-choice assessments are appropriate for assessing HOT skills.

Bacon (2003) explains that "multiple-choice tests have an economy of scale not found in constructed-response test." meaning that the time allocated for preparing, administering, and evaluating the test is not dependent on the number of students completing the assessment (p. 32). Kohn (2000) also notes the same economic advantage of multiple-choice formats. Both Kohn (2000) and Bacon (2003) note that multiple-choice assessments are widely used by highly regulated tests including Scholastic Aptitude Test (SAT), the American College Test (ACT), portions of several Advanced Placement (AP) tests, and most state mandated testing.

Bacon (2003) also identifies that multiple-choice questions require less time to answer, and therefore more questions can be answered in the same amount of time required to ask fewer constructed-response questions. This leads to "greater domain sampling" (p. 35). For example, if it takes a student three minutes to answer one free-response question, the same student can answer three or more multiple-choice questions on the same topic in the same amount of time. Each of the multiple-choice questions would be worded slightly differently, thereby providing a larger statistical sample and possibly ruling out interpretation errors associated with poorly worded questions. Several researchers (Bridgeman & Rock, 1993; Bridgeman & Lewis, 1994; Lukhele, Thissen, & Wainer, 1994; Bacon, 2003) have found the scores from multiple-choice sections of assessments (such as Advanced Placement (AP) test for biology and chemistry, and

Graduate Record Examination) to be comparable to the corresponding free-response section. For example, Bacon (2003) found that when a multiple-choice exam was held to the same standards as the short-answer exam the resulting scores were similar.

Another advantage to the multiple-choice assessment is that it is unambiguous to score. Multiple-choice answers are either correct or incorrect, making them comparable and consistent (Moss, 1992). Other traditional paper-and-pencil methods are more ambiguous to score; for example, Hasse et al. (2014) found that when evaluating the scoring competence of biology teachers, the number of possible answers to the free-response questions had to be limited so that the assessment could be reliably scored with a scoring rubric. Although used extensively within the United States, not many other countries use multiple-choice assessments (Kohn, 2000).

It is controversial as to whether multiple-choices format questions can assess HOT skills. Kohn (2000) depicts multiple-choice assessments as exercises that measure a student's ability to recite information crammed into short-term memory. The format of multiple-choice questions focuses only on whether or not the correct answer was selected and prohibits the student from demonstrating the reasoning behind how they selected the correct answer. "No matter how clever or tricky the questions are, multiple-choice tests simply do not measure the same cognitive skills as are measured by similar problems in free-response form" (Kohn, 2000, p. 8). Moss (1992) identifies the dependence of the educational system on multiple-choice assessments as having a negative impact on the

education system, claiming that the multiple-choice format is causing the curriculum to narrow in order to conform to the constraints of these tests.

The Debate over Performance-based Assessments

Performance-based assessments have been felt by many to be more beneficial in the long term to teaching and learning than are multiple-choice form assessments because they enable the evaluation of HOT skills (Moss, 1992; Harlen, 2000; Erdogan, Campbell, & Abd-Hamid, 2011; Puncochar & Klett, 2013; Harrison, 2014). Several educational research sources advocate for inquiry-based instruction that is assessed using performance-based assessments (Moss, 1992; Harlen, 2000; Erdogan, Campbell, & Abd-Hamid, 2011; Puncochar & Klett, 2013; Harrison, 2014).

Harrison's (2014) findings regarding the *Strategies for Assessment of Inquiry Learning in Science (SAILS)* was that the performance-based assessments (specifically, observing during class, reading homework questions, and watching learners interact in groups) produced extensive and descriptive assessment data. One of the advantages to performance-based assessment is that the student is not assessed once but continually as they work through the inquiry activity. The continuous assessment technique compensates for variations in student performance from day to day, allowing the teacher to have an accurate assessment of the learner's achievement potential.

Puncochar and Klett (2013) also employed performance-based assessments for evaluating the impact of inquiry. They found that randomly selecting of students' performance-based assessment artifacts (for example from a portfolio) were "critical" to

the accreditation process of universities (p. 49). There are concerns, however, that

evaluation systems focused on "reliability, efficiency, and comparability of assessments"

(Moss, 1992, p. 248) will favor multiple-choice assessments over performance-based

assessments.

There are also disadvantages to performance-based assessments. Performance-

based assessments have not been widely incorporated into assessments that determine the

federal funding and or scholarships, expectance to institutes to further the individuals'

education, or if an individual can acquire credit for an entire year of work, like the SAT,

ACT, and portions of several AP tests. As Harrison (2014) states, one of the reasons for

the slow incorporation of inquiry-based instruction into the classrooms of Europe is the

lack of inquiry-based science education questions on national and international

assessments. The absence of performance-based items on the national and international

assessments diminishes the confidence that parents and administration have in

performance-based formats (Harrison, 2014). Teachers are encouraged to prepare

students for state or district assessments that are in the multiple-choice format (NRC,

2000). Multiple-choice assessments are established (e.g. SAT, ACT, AP test etc.) where

as performance-based assessments are not commonly used in established formats for

national and international assessments (Harrison; 2014).

Another major limitation to performance-based assessment is that it requires one-

on-one interaction. For example, the SAILS project (Harrison, 2014) acknowledged the

limitations of performance-based assessment in that the teachers were not able to assess

every student during every inquiry activity. When assessing large masses of students, it is prohibitive to employ one-on-one interactions with every student (Puncochar & Klett, 2013).

There is also the issue of scoring performance-based assessments. Performance-based assessments give the students a degree of latitude in how they interpret and respond to the tasks that require "expert judgment to evaluate" (Moss, 1992, p. 230). Thus, the performance-based assessment's score can accidentally be biased by a scorer. For example to reduce scoring bias Erdogan, Campbell, & Abd-Hamid, (2011) employed 67 Turkish pre-service science teachers who worked in groups to evaluate one performance-based assessment designed to compare outcomes of performance-based instruction to teacher-based instruction techniques. The study found that many factors led to variance in the scores on the performance-based assessment.

In Puncochar and Klett's (2013) study, scoring bias was avoided by having the scoring rubric implemented by a committee of three volunteer university faculty members from diverse science backgrounds. It was found that the bias of one assessor could influence the score that the recipient received and that possible bias within the scoring system seems to only be mitigated by employing multiple scorers. A test that requires multiple scorers to interpret would not be applicable to the average high school biology classroom setting, however.

The debate between traditional paper-and-pencil and performance-based assessment is far from concluded. This report compares students' scores on three short

16

paper-and-pencil assessments to those a long performance-based assessment to determine the degree of correlation between the ranked scores. A strong correlation between the short paper-and-pencil assessments and the long performance-based assessment would indicate that the paper-and-pencil assessment is a good alternative to the performance-based assessment, thereby eliminating some of the identified problems associated with performance-based assessments. The following section looks at three different educational research projects that all assess HOT skills using paper-and-pencil method, yet each employs very different style of assessments.

Three Approaches to Assessing Higher-Order Thinking Skills

Myers and Burgess (2003), Hofstein et al. (2005) and Hohenshell and Hand (2006) developed assessments to evaluate student's HOT skills developed from inquiry-based teaching techniques. They each employed different assessment techniques consisting of a mixture of traditional and performance-based assessment techniques. This study drew from these three studies to develop three assessment tools. These three assessments were then compared against one another for equality and fairness, and applicability. Then the results were compared to the results from a semester long performance-based assessment to evaluate the validity of the assessments for assessing HOT skills.

Myers and Burgess (2003) Assessment

The assessment tool Myers and Burgess (2003) developed evaluated critical thinking and analysis skills associated with inquiry-based instruction. The students in

17

ASSESSING ASSESSMENTS

Myers and Burgess' (2003) study were enrolled in a redesigned intermediate-level organismal physiology inquiry-based laboratory course at the college level. The assessment tool, administered at the beginning and conclusion of the semester, consisted of a traditional paper-and-pencil assessment that included an experimental scenario (reading comprehension) with data and four free-response questions of increasing complexity. The four questions examined the students' ability to analyze data and experimental design. Myers and Burgess' (2003) assessment results showed an improvement in the critical thinking and inquiry skills of students enrolled in the inquiry-based laboratory course when compared to the students not taking the inquiry-based laboratory course.

Hofstein et al. (2005) Assessment

Hofstein et al. (2005) believed the aspects of inquiry-based instruction that develop long-term understanding of scientific topics and metacognitive skills related to a students' ability to ask questions. These researchers studied the ability of high school twelfth grade chemistry students to ask questions in general and inquiry-type questions in particular, and then translate their questioning skills into another learning situation, namely the critical reading of a scientific article and a traditional paper-and-pencil assessment. This was accomplished by exposing students to a novel experiment where they were instructed to first generate inquiry questions and then determine which questions they wanted to evaluate further. The questions generated by the students were analyzed in three ways: (a) the number of questions generated (b) the quality of the

18

questions generated HOT or LOT questions, and (c) the quality of the questions that were selected by the student for further investigation. Hofstein et al. (2005) stated in their final discussion that the "student's ability to ask questions should be seen as an important component of science literacy and should not be overlooked" (p. 802).

Hohenshell and Hand (2006) Assessment

Hohenshell and Hand's (2006) evaluation tool measured students' knowledge and understanding developed as a result of their inquiry-based instruction. The study population was ninth and tenth grade high school students that studied cellular physiology over seven weeks as part of their Biology class. These researchers evaluated the efficiency of a writing-to-learn technique, which they called "Science Writing Heuristic" (SWH). SWH is a semi-structured inquiry learning experience where high school students (a) identify patterns in data, (b) construct and support knowledge based on that data and prior knowledge, (c) make connections between data, claims and evidence, and (d) develop scientific arguments.

The SWH technique that incorporated inquiry and scientific arguments was compared to the traditional formal laboratory written report technique. Hohenshell and Hand's (2006) assessment consisted of a pre-test, a first-post-test at the completion of the instructional phase of the work, followed by a guided writing activity, and a second-post content test. The tests consisted of a traditional paper-and-pencil method that included ten multiple-choice questions measuring LOT recall knowledge and four free-response items measuring HOT conceptual knowledge. The content tests were accompanied by two

performance-based assessments: an open-ended survey and student interviews. The tests analyzed the students' science-based skills in the content area, whereas the survey and interviews analyzed the students' experience of the process of science, indicating the development of metacognitive awareness.

Summary

The assessments created by Myers and Burgess (2003), Hofstein et al. (2005) and Hohenshell and Hand (2006) were designed to evaluate HOT as a product of inquiry-based instructional techniques, and each assessment is very different. Considering the various assessment techniques employed by educational researchers and approaches that are debated by educational researchers, I have developed three assessment tools (Assessment 1, Assessment 2, and Assessment 3) that are loosely based on the assessments developed by Myers and Burgess (2003), Hofstein et al. (2005), and Hohenshell and Hand (2006) for the evaluation of HOT as a product of inquiry-based instruction in my classroom. These will be discussed in the following chapter. The intent of this study is to determine if any of these assessments can be used as a self-evaluation tool to evaluate the effectiveness of inquiry-based teaching techniques with respect to developing HOT skills in students.

Chapter 3 Methodology

Description of the Participants

There were approximately two hundred tenth grade biology students within the high school where the study took place during the 2015-2016 school year. The majority of the student population was of low socioeconomic standing. All of the student body was subsidized for both lunch and breakfast programs. The tenth grade students were broken into ten biology classes taught by three biology teachers. Four biology classes, one of which had a co-teacher to facilitate special education students within that class, participated in the study. The four participating biology classes (n=85) were instructed by the same instructor, who was also the researcher. All four participating biology classes participated in three inquiry units prior to completing the assessments designed for this study.

Inquiry Lessons

Four classes of tenth grade high school biology students, as part of the course curriculum, participated in three inquiry units to develop HOT skills. The first two inquiry units were guided, meaning the students were presented with a question and instructed to design an experiment to address the question. The two guided inquiry units were designed to engage the students in scientific content related to osmosis and DNA extraction, respectively; both units can be found attached in the appendix (See Appendix A and B). The students designed their experiments, conducted their experiments,

formulated explanations based on their experiments, and communicated their results for both units during the second and third marking period of the 2015-2016 school year.

The third inquiry unit involved forest ecology in a semester-long open inquiry project took place during the fourth and final marking period for the 2015-2016 school year. Open inquiry involves the students generating their own question to investigate rather than having questions presented to them. The high school biology classes developed scientifically-oriented questions related to forest ecology, collected evidence, used evidence to evaluate explanations, formulated explanations from the evidence they collected or researched, evaluated and justified their explanations, and communicated their results. The Final assessment scoring rubric, used to evaluate the open inquiry forest ecology unit, is Appendix D and the open inquiry forest ecology unit is Appendix C.

Assessments and Scoring

Three assessment tools (Assessment 1, Assessment 2, and Assessment 3) were developed for this study that were loosely based on the assessments developed by Myers and Burgess (2003), Hofstein et al. (2005), and Hohenshell and Hand (2006), respectively, for the evaluation of HOT as a product of inquiry-based instruction. See Table 2 for the exact dates during the fourth marking period that each assessment was administered.

Table 2

Assessments Administered by Class and Date

	Class 1	Class 2	Class 3	Class 4
Assessment 1	Did not take	5/24/2016	5/23/2016	5/23/2016
Assessment 2	5/24/2016	5/23/2016	Did not take	5/24/2016
Assessment 3	5/23/2016	Did not take	5/24/2016	Did not take

The three assessments were compared against one another to determine (a) equality and fairness and (b) system applicability. Then the three shorter assessments were compared against a longer performance-based assessment to determine the degree of comparison for the evaluating HOT skills. The three shorter assessments were altered from those prepared by Myers and Burgess (2003), Hofstein et al (2005) and Hohenshell and Hand (2006) in the following manner.

Multiple-choice questions were substituted for free-response questions in a few of the assessments to increase consistency between the three assessments so that each assessment contained multiple-choice questions. This was done because multiple-choice questions increase the expedience of grading and reduce ambiguity when scored by one individual. The high school biology population who volunteered to participate in this study verbally communicated a level of comfort associated with taking multiple-choice type questions and expressed a resistance to free-response questions. Additionally multiple-choice questions are widely employed by highly regulated standardized tests and have been for most of the educational career of the tested population. A practice

opportunity for the students was also created by assimilating the multiple-choice format that is similar to the state mandated standardized test. The opportunity to practice taking tests similar to the standardized test was encouraged by school administration. The implications of such a change will be further discussed in the conclusion/discussion section of this report.

Myers and Burgess (2003) Assessment

Assessment 1 was loosely based off of the assessment Myers and Burgess (2003) developed for evaluation of a redesigned intermediate-level organismal physiology inquiry-based laboratory college level course. Myers and Burgess' (2003) assessment consisted of four traditional paper-and-pencil free-response questions related to an experimental scenario and supporting data. The free-response questions increased in complexity, asking students to (a) create a graph, (b) describe the nature of the relationship between two aspects of the experimental results, (c) interpret and synthesize a response to the experiment and data, and (d) explain how to improve the experiment. Myers and Burgess (2003) implemented the assessment tool to compare pre and post scores of students in a control and experimental group. Myers and Burgess' (2003) assessment was scored on a scale from 0 to 3 points. The individual questions were assessed independently and improvement scores were calculated.

Development of Assessment 1

For the Myers and Burgess (2003) study the testing population included college level intermediate-level organismal physiology laboratory student, whereas the testing

24

population in this study was composed of tenth grade high school biology students. Due to the difference in the testing population, the topic of the experimental scenario was adjusted from Myers and Burgess' (2003) "Investigating the Effect of Step Cadence on Heart Rate" to the scenario used in Assessment 1: "The Effect of Plowing on Earthworm Populations."

Table 2 is a comparison between the assessment tool Meyers and Burgess (2003) used and Assessment 1 prepared for this study. The time restraints prohibited the implementation of a pre- and post-test. The pre- and post-evaluations are employed to evaluating an improvement in HOT skills as a product of the inquiry instruction techniques (Appendix A, B, and C). The omission of the pre- and post-evaluations only affects a measure of improvement, which is not the focus of this report. The intent of this report was to identify an assessment that can be used by a teacher as a self-evaluation tool would allow them to evaluate their inquiry teaching technique with respect to HOT skills. Once that assessment is identified then in practice it could be administered before and after an inquiry lesson to determine the development of HOT skills in the students.

This study did not use a control group. A control group is not required because the focus of this report is not to determine the impact of a specific inquiry technique. The intent of this report is to identify a short assessment that could be used rather than a long performance-based assessment to evaluate HOT skills. For this study the semester long final project performance-based assessment was compared against each of the three assessments rather than comparing them to a control group.

25

ASSESSING ASSESSMENTS

The most significant difference between the original research assessments and the assessments created for this study, is that the free-response questions were replaced by multiple-choice questions. This affected the Assessment 1 by reducing the point scale. The point scale of Assessment 1 was reduced from a 12-point grading scale that was used on the Meyers and Burgess (2003) assessment, to a 4-point grading scale (See Table 3 above).

Table 3

Comparison between Meyers and Burgess Assessment and Assessment 1

	Meyers and Burgess (2005)	Assessment 1
Type of Questions	Free-Response	Multiple-Choice
Number of HOT Questions	2	2
Number of LOT Questions	2	2
Pre and Post Evaluation	Yes	No
Control Group	Yes	No
Grading Scale	12 Points	4 Points

The four questions were also converted into multiple-choice questions for reasons discussed previously. Due to time restraints no pre-test was administered in this study, preventing improvement score from being calculated. Table 4 provides a comparison of the Myers and Burgess (2003) questions and the questions included in Assessment 1. Note that in Assessment 1, students were provided an experimental scenario and a graph to answer the questions. Assessment 1 as presented to the students in the study can be found as Appendix E to this document.

Table 4

Comparison of Myers and Burgess Assessment and Assessment 1 Questions

	Myers and Burgess (2003) Questions	Assessment 1	Bloom's Taxonomy Level
Question 1	Create a graph to help you visualize the effect of step cadence on heart rate.	What question did the scientists who collected this data want to answer?	Knowledge
Question 2	What is the nature of the relationship between step cadence and heart rate?	Where and when were the most earthworms found?	Comprehension
Question 3	On the basis of your analysis, what advice would you give Taliz about the step cadences she should use in her routines if she wants to keep the heart rate of her students between 110 and 120 beats/min during the 45 min class?	What does the data in the graph show?	Application
Question 4	If you wanted to predict heart rate from step cadence with great precision, how would you improve the experimental design?	A scientist who wants to study the effects of a new fertilizer on plants would probably	Synthesis

Assessment 1 was loosely based off of the Myers and Burgess (2003) assessment used in their Inquiry-based Laboratory Course Improves Students' Ability to Design Experiments and Interpret Data Article Published by The American Physiological Society.

Two of the questions on Assessment 1, Question 3 and Question 4, are HOT

questions. For the third question, the student is asked to identify the relationship between

variables in a given set of data by asking "What does the data in the graph show?" The

student has to select from a list of possible conclusions by applying the data presented in the table. The possible answer selections were (a) unplowed soil has more earthworms than plowed soil, (b) plowed soil has more earthworms than unplowed soil, (c) plowing of soil has no effect on the number of earthworms, and (d) the number of earthworms cannot be predicted.

The fourth question required students to choose an experimental design in response to the prompt, "A scientist who wants to study the effects of a new fertilizer on plants would probably…" This is followed by a list of possible experimental designs that would satisfy the requirements stated in the question: (a) give each experimental group the same amount of fertilizer, (b) not worry about measuring the amount of fertilizer used, (c) include a control group that received no fertilizer, or (d) use different numbers of plants in each group.

The four multiple-choice questions on this assessment were scored for correctness, with students receiving a point for a correct answer and no points for an incorrect answer. Thus, each student received a score from zero to four.

Hofstein et al (2005) Assessment

Assessment 2 consisted of a performance-based practical portion and a paper-and-pencil critical reading component. This assessment was loosely based on the Hofstein et al. (2005) assessment and was administered to a high school chemistry class, in which students experienced a performance-based practical test and a paper-and-pencil critical reading component. The performance-based practical portion involved the students

performing a new experiment involving a chemical reaction consisting of two unknown substances. The experiment was followed by a practical questionnaire consisting of five free-response questions. The second part was a paper-and-pencil portion consisting of a critical reading article related to an unknown chemical reaction accompanied by eight free-response questions. Hofstein et al. (2005) analyzed only two of the eight free-response critical reading questions: (a) "Write down all the questions that you would like to ask after reading this article" and (b) "From the list of questions, select the most interesting one that you would like to investigate" (p. 797). The responses of the control and experimental groups were compared. The results of the both the practical section and the critical reading section were both evaluated (a) for the number of questions each student presented, (b) for the level of the questions presented, and (c) for the level of the question chosen to investigate further.

The levels of the questions generated in both the practical and the critical reading portions were evaluated to be "low or high-level type questions" (p. 797) based on their content. For the practical, LOT questions were questions related to "facts and explanations of the phenomena that were observed in the experiment performed by the students" (p. 797). For the critical reading portion LOT questions were those that were textual questions based on information found within the text. For both the practical and the critical reading portions, HOT type questions require further investigation to answer, either through further experimentation or literary research.

29

Development of Assessment 2

Due to the test population consisting of high school biology students, the topic of the novel experiment (practical test) in Assessment 2 involved a biologic anomaly involving rods and cones of the retina of the human eye. The novel experiment associated with the practical portion was accompanied by a free-response portion identical to the one Hofstein et al. (2005) employed. The topic of my critical reading article involved "How we see in color" (Copper & Kazilek, 2016). See Appendix F for Assessment 2.

The critical reading component was accompanied by six multiple-choice questions of increasing complexity. The critical reading question set was changed from free-response questions to multiple-choice format: to increase the consistency between the three assessments, to reduce ambiguity and the efficiency of the grading process, to placate volunteer participation, and at the recommendation of administration to provide an opportunity to practice for standardized testing.

Table 5 is a comparison between Hofstein et al (2005) assessment and Assessment 2 prepared for this study. As shown in Table 5, Hofstein et al (2005) used a control group whereas Assessment 2 did not have a control group. Hofstein et al (2005) did not score their assessment rather they compared the number of questions, and the types of questions asked by their test population to that of the control population. Another difference between the Hofstein et al (2005) assessment and Assessment 2 is the inclusion of more LOT questions in Assessment 2 then in Hofstein et al (2005) assessment.

30

Table 5

Comparison between Hofstein et al. (2005) Assessment and Assessment 2

	Hofstein et al.	Assessment 2
Practical Portion	Yes	Yes
Type of Questions	Free-Response	Free-Response
Number of HOT Questions	3	3
Number of LOT Questions	0	0
Critical Reading Portion	Yes	Yes
Type of Questions	Free-Response	Multiple-Choice
Number of HOT Questions	2	2
Number of LOT Questions	0	4
Pre and Post Evaluation	No	No
Control Group	Yes	No
Grading Scale	Not Scored	23 Points

The first four multiple-choice questions were LOT questions, the answers to which could be found within the text. An example is the question, "Having a genetic defect that prevents someone from having all three types of cone cells would result in…" The answer to this question is color blindness and was located within the text of the article. The last two questions were HOT questions because they required the application and synthesis of the information in the article. For example, one question was, "An animal that is active primarily at night such as a lion would have more of what type of photoreceptor cell?" To obtain the correct answer the student needed to apply the information from the article to the new scenario about a lion. The critical reading multiple-choice questions were scored for correctness receiving a point for a correct answer and no points for an incorrect answer.

The practical portion was evaluated for the number of questions each student presented, the level of the questions, and the level of the question chosen to investigate further. The levels of the questions were evaluated as low or high-level questions based on the content. LOT questions were questions that had one simple answer that could be answered with minimum research. An example is the question, "Why did I see a white circle when nothing was there?" This question is answered by reading the introduction to the experiment.

HOT questions were questions that could be answered through further experimentation. These questions required the student to put ideas together to construct a new experiment. An example of a HOT question is, "Would the circle have to be red?" This is considered a HOT question because the student could then conduct an experiment with a different color circle to determine the effect of color on overloading of the photoreceptors. LOT questions received one point and HOT questions received two points. See Table 6 for more examples of LOT and HOT questions that the participants wrote and then selected as questions they would like to investigate further.

Table 6

Examples of Questions Chosen by the Participants for Further Investigation

Low Order Thinking Skill Questions (1 point)	Why is it a brighter white than the card?
	What am I suppose to see?
	Why wasn't we aloud to blink?
	Why did I see a white circle when nothing was there?
	Why was there a glow after 60 seconds?
	Why did I have to stare at that?
	Why did I see the blue dot where the red one was?
	No answer
Higher-order Thinking Skill Questions (2 points)	Would I see a blue dot on the back if the paper was black?
	Do different color dots change colors when you stare at it and then flip it over?
	Would the circle have to be red?
	Does it work with different colors?
	Does the color of the original dot affect the color of the fake one?
	If stared longer, would more green dots appear?
	Does everyone see the dot? (Colorblind people too)
	Does eye color affect what people see when they look at the dot?

To give the reader a sense of the scoring for Assessment 2, consider Student 303. This student posed five questions (5 points): two of the questions were LOT questions (2x1=2 points) and three of the questions were HOT questions (3x2=6 points). The question that was selected for further investigation "Would I see a blue dot on the back if the paper was black?" was a HOT question (2 points) because this questions could lead to further experimentation of repeating the test with a black paper rather than a white paper. Student 303 also answered all 6 multiple choice questions correctly (6 points). Student 303 received a composite score of 21 points. See Table 7 for a summary of this scoring as

well as that of Student 308. The question Student 308 selected, "Why is it a brighter white than the card?" is considered a LOT questions because it could be answered from reading the introduction to the experiment: "[T]he exhausted red cones will not transmit an impulse to the brain. But the blue-sensitive and green-sensitive cones in the area will send signals to the brain. Thus, the brain will be tricked into perceiving a blue-green (or brighter white) dot where one doesn't exist."

Table 7

Example of Composite Score Calculation for Assessment 2

Student	Number of Questions (0-5 points)	Number of LOT Questions	Points for LOT Questions	Number of HOT Questions	Points for HOT Questions
303	5	2	2	3	6
308	1	1	1	0	0

Student	Question Chosen			Level of Chosen Question	Points for Question Chosen (1-2)
303	Would I see a blue dot on the back if the paper was black?			High	2
308	Why is it a brighter white than the card?			Low	1

Student	Number of Multiple Chose Correct	Students Composite Score (out of 23 points)
303	6	21
308	5	8

Hohenshell and Hand (2006) Assessment

The third assessment developed for this study was loosely based on that developed by Hohenshell and Hand (2006). Their assessment consisted of (a) ten

multiple-choice LOT recall knowledge questions, (b) four free-response HOT conceptual knowledge questions, (c) an open-ended survey question, and (d) an interview of a small portion of the experimental population on the topic of cell physiology. The test population consisted of mixed classes of ninth and tenth grade students in an advanced placement biology course covering a seven-week unit on cell physiology.

Development of Assessment 3

Assessment 3 for this study consisted of ten multiple-choice LOT recall knowledge questions, and contained ten, rather than four, free-response HOT conceptual knowledge questions. No interviews were conducted due to time restraints during the testing period. The content of the two tests was different; Hohenshell and Hand's (2006) assessment was about cellular physiology and the content of Assessment 3 was forest ecology. See Appendix G for Assessment 3.

Table 8 is a comparison between Hohenshell and Hand (2006) assessment to that of Assessment 3. The difference between the Hohenshell and Hand (2006) assessment and Assessment 3 is an increase in the number of HOT free-response questions.

Table 8

Comparison between Hohenshell and Hand (2006) Assessment and Assessment 3

Type of Questions	Hohenshell and Hand Free-Response	Assessment 3 Free-Response
Number of HOT Questions	4	10
Number of LOT Questions	0	0
Type of Questions	Multiple-Choice	Multiple-Choice
Number of HOT Questions	0	0
Number of LOT Questions	10	10
Pre and Post Evaluation	Yes	No
Control Group	No	No
Grading Scale	14 Points	30 Points

The multiple-choice LOT recall questions were scored for correctness receiving a point for a correct answer and no points for an incorrect or no answer. The free-response HOT conceptual knowledge questions were given two points if the response demonstrated the application of knowledge to the situation presented. Responses that demonstrated knowledge of the meaning or the definition without applying that meaning to the situation presented received one point, and responses that were completely wrong or absent did not receive any points. See Table 9 for examples of the grading scale for the free-response questions for Assessment 3.

Table 9

Examples of Free-response Answers at Different Levels on Assessment 3

Directions: List in complete sentences an example of the following that you observed in the forest ecosystem.	
Question	A species you observed:
Examples of free-response answers that demonstrated HOT (2 Points)	A rhododendron. A brown squirrel. I observed an oak tree.
Examples of free-response answers that demonstrated LOT (1 Points)	Of plant and insect. I observed a bird. A group they can make offspring.
18% of the participants did not answer the questions. (0 Points)	

Answers to the question "List in complete sentences an example of a species you observed in the forest ecosystem." that demonstrated recall of definition without applying it to the forest ecosystem were considered LOT questions. For example, the answer "A group they can make offspring" is similar to the definition of a species (A group of living organisms consisting of similar individuals capable of producing offspring) and does not apply that definition to the forest ecosystem therefore, it is a LOT answer and received one point.

Table 10 provides a detailed example of how Assessment 3 was scored for Student 303 (Assessment 3 questions can be found in Appendix G).

Table 10

Sample Scoring Calculation for Assessment 3

Multiple Choice Section		
Student	Number Correct (0-10 points)	Total Points
303	6	6
Free-Response Section		0-2 points
#11	The sun feeds the plants, the plants feed the but, the but feeds the bird.	2
#12	Of plant and inceact.	0
#13	Seen no wildlife	1
#14	The tree give many a home.	0
#15	The reproduction and living of the animals.	0
#16	Death and dieing out of a bread.	0
#17	The bugs eat off the tree the birds live in the tree get the bugs.	2
#18	Room to grow.	2
#19	The plant gives many energy to pass throw.	2
#20	The leaves, then the bugs, then the birds.	2
	Total Points	11
Assessment 3 Score for Student 303		17

Final Project Score

The Biology students participated in a semester-long open-inquiry unit involving forest ecology during the fourth and final marking period of the 2015-2016 school year. The students developed scientifically-oriented questions related to forest ecology, collected evidence, used evidence to evaluate explanations, formulated explanations from the evidence they collected or researched, evaluated and justified their explanations, and communicated their results. The inquiry projects were graded with the scoring rubric

found in Appendix D. The scoring rubric can be broken up into two different sections: the

Inquiry Participation section and the Inquiry section. The scoring rubric was obtained

from a chemistry teacher with over thirty years of experience who has used it to score

inquiry for more than three years. The scoring rubric had been altered from an inquiry

scoring rubric published in The Science Teacher (Grady, 2010).

Only the Inquiry Section scores were used in this analysis. Scores were assigned

based on the students' performance as Smashing/Excellent receiving all the possible

points for the category, Polished/Pretty Darn Good receiving most of the possible points

for the category, Intact/Pretty Good receiving some of the point for the category, and

Unfortunate receiving none of the points for the category. The Inquiry section had 135

points possible divided between eight sections. The categories ranged in point from 35 to

10 possible points. The points acquired for each section were added to calculate the final

composite score. The categories were (a) Make Observations (20 points), (b) Define

Questions (10 points), (c) Quality of your research (35 points), (d) Communication (10

points), (e) Conclusion Claims and Evidence (20 points), (f) Quality of your Presentation

(20 points), (g) Consider New Evidence (10 points), and (g) Add to Explanation (10

points). The sections of the final project scoring rubric listed above coincide with the

NRC 6 Fundamental Abilities of Inquiry (Barrow, 2006). Table 11 shows how the

sections of the final project scoring rubric and the 6 Fundamental Abilities of Inquiry

align. There is a one to one alignment with the exception of the sixth ability of inquiry:

communicate and defend a scientific argument. This inquiry ability is addressed in three

sections of the final project scoring rubric: (a) Communication, (b) Conclusion Claims and Evidence, and (c) Quality of your presentation. Students were evaluated by the instructor at the time they presented their final project.

Table 11

Comparison of the NRC's six Fundamental Abilities of Inquiry and the Final Project Scoring Rubric

	NRC's six fundamental abilities of inquiry	Sections of Final Project Scoring Rubric
1	Identify questions and concepts that guide investigations (student formulate and testable hypothesis and an appropriate design to be used);	Define Questions
2	Design and conduct scientific investigations (using major concepts, proper equipment, safety precautions, use of technologies, etc., where students must use evidence, apply logic, and construct an argument for their proposed explanations);	Quality of you Research
3	Use appropriate technologies and mathematics to improve investigations and communications;	Make Observations
4	Formulate and revise scientific explanations and models using logic and evidence (the students' inquiry should result in an explanation or a model);	Consider New Evidence
5	Recognize and analyze alternative explanations and models (reviewing current scientific understanding and evidence to determine which explanations of the model is best);	Add to Explanation

Table 11 (cont.)

NRC's six fundamental abilities of inquiry	Sections of Final Project Scoring Rubric
6 Communicate and defend a scientific argument (students should refine their skills by presenting written and oral presentations that involve responding appropriately to critical comments from peers).	Communication Conclusion Claims and Evidence Quality of your Presentation

Data Collection and Analysis

Ethics

Prior to data collection, a passive consent form was sent home to the parents/guardians of each of the students. Scores of students that did not consent were excluded from the study. Anonymity was maintained for the students by assigning a unique identification number to each student's data that was coded on the open inquiry project and on the correlating assessments.

Data Collection

In the four biology classes there were 83 students enrolled during the time of the final project and the administration of the assessments. Only 45 students participated in the final project, the remaining students did not complete the final project due to poor attendance or because they preferred to receive zero points for the assignment.

All four participating biology classes participated in three inquiry units prior to completing two of the three assessments designed for this study. The assessments were assigned randomly to the four classes. The first class participated in Assessment 2 and

Assessment 3. The second class participated in Assessment 1 and Assessment 2. The third class participated in Assessment 1 and Assessment 3, and the fourth class participated in Assessment 2 and Assessment 1 (see Table 2 for a summary).

Each of the three assessments were only administered once not allowing for make-up days due to absences. There were 29 students who completed both the final project and Assessment 1. There were 32 students who completed both the final project and Assessment 2. There were 28 students who completed for both the final project and Assessment 3.

Comparison Analysis: Evaluation of Higher-Order Thinking Skills

After all three assessments (Assessment 1, Assessment 2, and Assessment 3) and the final inquiry projects were scored as described above, the scores on each assessment were compared statistically to the final project score using the Spearman rank-order correlation coefficient, to determine whether there was a correlation between the two respective sets of scores. The formula for calculating the Spearman rank-order coefficient rho is given in Figure 1.

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} D_i^2}{n(n^2 - 1)}$$

Figure 1. Formula for Spearman rank-order correlation coefficient rho.

ASSESSING ASSESSMENTS

For example, the composite scores from Assessment 3 were compared to the final inquiry project score and a Spearman rank-order correlation coefficient rho was calculated. This was accomplished by first ranking the student scores on each assessment. The student that received the highest score received a one and the student that received the second highest score received a two and so forth until all of the students that participated in both Assessment 3 and the final project were ranked. If there were more than one student that received the same score the scores were considered to be tied. Tied ranks were calculated by assigning all the tied students with an average of the ranks that would have been designated to the students. For example, for Assessment 3 there were two students that received a score of 26. They would have been designated to receive rankings of 2 and 3, but because they tied they both received a ranking of the average of these rankings, 2.5. The difference (D) in the rankings of the two assessments, Assessment 3 and the final project for example, was calculated by subtracting a student's ranking on one assessment from their ranking on the other. This is the value of D and n is the sample size in the formula in Figure 1. Spearman correlation coefficient rank of 0 to 0.19 indicate a very weak degree of correlation, 0.20 to 0.39 indicate a weak degree of correlation, a rank of 0.40 to 0.59 indicate a moderate degree of correlation, 0.60 to 0.79 indicate a strong degree of correlation and 0.80 to 1 indicate a very strong degree of correlation between the two assessments.

For Assessment 1 and Assessment 2, there were a large number of tied rankings, more than half the sample population were tied. Therefore, on Assessment 1 and

Assessment 2 the Spearman rank-order correlation coefficient correction factor was accounted for in the calculations. The number of groupings of tied ranks (g) was determined. For Assessment 1 there were four groups of tied ranks because more than one student received each score of 4, 3, 2, and 1 therefore there were four groups that had tied ranks (g = 4).

$$T_x = \sum_{i\text{-}t}^{g} (t_i^3 - t_t)$$

Figure 2. The tied score correction value equation.

The sum of squares tied ranking correction factor was calculated based on the number of tied groups. The sum of square tied ranks was then corrected using the sum of squares tied ranking correction factor. For example, there were nine students that received a score of four on Assessment 1. Nine was squared, and then nine was subtracted from that product. This was done for each tied score and then these products were added from Assessment 1 to calculate the correction value (T_y). See Figure 2 above for the equation to calculate T_y and T_x. The same was done for the final project scores (T_x). These corrected values were used to calculate the Spearman rank-order correlation coefficient (Figure 3).

$$r_s = \frac{-6Sd^2 - (T_x + T_y)/2}{\sqrt{(N^3 - N) - (T_x + T_y)(N^3 - N) + T_x T_y}}$$

Figure 3. Spearman rank-order correlation coefficient correction factor equation.

The p-value was also calculated to determine the significance of the results. The hypothesis was that there is a correlation between the two respective assessments (Assessment 1 and the final project, for example) and the null-hypothesis being that there is not a correlation between the two respective assessments. A p-value less than 0.05 would indicate strong significance that the null-hypothesis (suggesting that there is not a correlation between the two assessments) is invalid. A p-value greater than 0.05 indicates a weak significance for the alternative hypothesis (i.e., a correlation between the two assessments) and the null-hypothesis cannot be rejected.

Analysis of Equality and Fairness

The content and the format of the assessments were compared to see if the assessments provided an equal access to resources and if prior knowledge creating a bias. The required resources for all three assessments involved reading assistance because no other materials were required, such as a calculator or graph paper. The three assessments were also compared against one another to see if questions could be answered from the information provided within the assessment (such as within the critical reading section) or the questions required prior knowledge about a specific content area.

Analysis of System Applicability

An assessment that has adequate system applicability would be an assessment that does not require more than one class period (55 minutes) to administer, is easy to interpret, and requires minimal resources. The assessments were compared against one another for the amount of consumable materials required, and the time each required for interpretations (grading). The time it took to interpret each assessment was calculated by recording the time it took to grade ten student assessments. Then the time was divided by ten to get the average time per paper required to interpret each assessment.

Chapter 4 Results

The intent of this study was to find an assessment that can be used as a self-evaluation tool to evaluate the effectiveness of inquiry-based teaching techniques with respect to developing HOT skills in students. The research question was: Of the three Assessments (Assessment 1, Assessment 2, and Assessment 3) which demonstrate the following characteristics: (a) ability to evaluate HOT skills, (b) equality and fairness, and (c) system applicability?

Evaluate Higher-Order Thinking Skills

A comparison of the composite scores from each of the three assessments separately to the inquiry score of the open inquiry final project was made with the Spearman rank-order correlation coefficient appropriate to the sample data.

A Spearman's correlation was first run to determine the relationship between the scores of the 29 students that participated in both the final inquiry project and Assessment 1. There was a weak positive correlation found between the score on the final project and the score on Assessment 1 ($r_s = 0.39$, n = 29, p = 0.06) (see Table 12). The resulting score is on the cut off between indicating a weak (0.39) and a moderate (0.40) level of correlation. The p-value is slightly greater than 0.05 indicating a weak significance for the hypothesis, that there is a correlation between the Assessment 1 and the Final Project scores.

Table 12

Comparison of Assessment 1 and the Final Project Composite Scores

	Final Project	Assessment 1
Total Score Possible	135	4
Mean	91.8	3
Standard Deviation	27.9	0.89
Comparative Analysis of the Final Project Score and Assessment 1 Score		
n	29	
Spearman coefficient	0.39*	
p-value	0.06**	

*For Spearman rho a strong correlation is between 0.7 and 1.

**p-value 0.05 is the cut off for significance

For Assessment 3, Spearman's correlation was run on the scores of 28 students. This also showed a weak positive correlation between the score on the final project and the score on Assessment 3 ($r_s = 0.35$, n = 32, p = 0.17) (see Table 13). The p-value is the largest between the three assessments indicating a strong significance that for the null-hypothesis, that there is not a correlation between the Assessment 3 and the Final Project scores.

Table 13

Comparison between Assessment 3 and the Final Project Composite Score

	Final Project	Assessment 3
Total Score Possible	135	30
Mean	93.9	18.2
Standard Deviation	28.2	5.50

Comparative Analysis of the Final Project Score and Assessment 3 Score

n	28
Spearman coefficient rho	0.35*
p-value	0.17**

*For Spearman rho a strong correlation is between 0.7 and 1.

**p-value 0.05 is the cut off for significance

Finally, Assessment 2 had the weakest positive correlation between the scores of 32 students on the final project and on Assessment 2 ($r_s = 0.31$, n = 28, p = 0.06) (see Table 14). The p-value is slightly greater than 0.05 indicating a weak significance for the hypothesis, that there is a correlation between the Assessment 2 and the Final Project scores. Thus the null-hypothesis, that there is not a correlation between Assessment 2 and the Final Project score, cannot be rejected.

Table 14

Comparison between Assessment 2 and the Final Project Composite Score

	Final Project	Assessment 2
Total Score Possible	135	23
Mean	86.2	11.9
Standard Deviation	28.4	4.07

Comparative analysis of the Final Project Score and Assessment 2 Score

n	28
Spearman coefficient gamma	0.31*
p-value	0.06**

*For Spearman rho a strong correlation is between 0.7 and 1.

**p-value 0.05 is the cut off for significance

Equality and Fairness

The content and the format of the assessments were compared to see if the assessments provided an equal access to resources and if prior knowledge creating a bias.

Assessment 1 consists of a short article accompanied by a graph of data. Assessment 1 is equally accessible to the students because the article was read aloud to the entire class. The multiple choice questions were read to students with individual education plans (IEP) that required reading assistance. Prior knowledge was not a factor with this assessment as the questions were based off the information obtained from the reading and the graph. For example, the answer to question number two, "Where and when were the most earthworms found?" could be interpreted from the provided bar graph "Unplowed soil, Spring 1995."

ASSESSING ASSESSMENTS

Like Assessment 1, Assessment 2 had equal access to resources because it was administered by reading the directions and the article aloud to the entire class. Points were not deducted for spelling or grammar errors so that there was not a bias. This assessment is based on the pretext that the students did not have prior knowledge of the experiment. The critical reading article explanation of the phenomenon they experience in the experiment is intended to be novel. Both Assessment 1 and Assessment 2 have the same degree of equality and fairness.

Assessment 3 required prior knowledge of ecology, specifically forest ecology that was part of the open inquiry lesson taught prior to taking the assessment. Additional knowledge obtained outside of the classroom curriculum was not part of this assessment. This assessment was read in its entirety to any students requiring this as part of their IEP, reading assistance making it equally accessible to all students.

As depicted in Table 15, the only assessment that requires prior knowledge is Assessment 3, otherwise all three assessments appear to be uniformly equal and fair.

Table 15

Comparative Analysis of Equality and Fairness of the Three Assessments

	Accessible to all Students	Required Prior Knowledge
Assessment 1	Yes	No
Assessment 2	Yes	No
Assessment 3	Yes	Yes, but taught as part of the curriculum

System Applicability

An assessment that has adequate system applicability would be an assessment that does not require more than one class period (55 minutes) to administer, is easy to interpret, and requires minimal resources. None of the three assessments exceeded one class period to administer.

Table 16 is a comparison between the three assessment of the time required to interpret the assessments. The time it took to grade ten student assessments was recorded and then averaged to determine the time per paper required to interpret each assessment. Assessment 3 requires the most time per paper to interpret followed by Assessment 2. Assessment 1 requires the least amount of time to interpret.

Table 16

Comparative Analysis of Time Required to Grade each Assessment

	Time to grade 10 papers (minutes)	Time per paper (minutes)
Assessment 1	0.42	0.04
Assessment 2	9.43	0.94
Assessment 3	13.1	1.31

Finally, to be applicable the assessment must require a minimal amount of resources. Assessment 1 required one class set of the critical reading article and graphical data, which took up one page of paper. The class set of this article could be re-used for multiple groups of students as long as they are instructed to refrain from writing on the class set. The multiple-choice question set was printed on a half sheet of paper. The students recorded their answers on this paper. Therefore, Assessment 1 required a half

sheet of paper to be consumable, with a class set of one page each that can be re-used for assessing multiple groups.

Assessment 2 required an index cards with red dots that had to be assembled prior to the assessment. This class set of index cards was required but could be re-used for multiple groups of students. The class set of instructions for the practical portion, the critical reading article and the multiple choice section could also be re-used for multiple groups of students as long as the students refrain from writing on them. The consumable material for this assessment consists of one sheet of lined paper for students to record their multiple-choice answers and practical responses.

Assessment 3 required one sheet of paper printed on both-sides, one side with the ten multiple choice questions and the alternate side with the free-response questions. This paper was consumed as students recorded their responses on it. Table 17 is a comparison of the consumable materials for the three assessments.

Table 17

Comparison of Consumable Materials for the Three Assessments

	Consumable Materials Per Assessment
Assessment 1	1/2 sheet of copy paper
Assessment 2	1 sheet of line paper
Assessment 3	1 sheet of copy paper

Chapter 5 Discussion and Conclusions

The intent of this study was to identify an assessment that can be used as a self-evaluation tool to evaluate the effectiveness of inquiry-based teaching techniques with respect to developing HOT skills in students. The three assessments were evaluated for (a) equality and fairness, (b) system applicability and (c) their ability to evaluate HOT skills as defined by Bloom's taxonomy. Table 18 below is a comparison between the three assessments and the criteria set to evaluate them.

Table 18

Comparison of the Three Assessments for Equality, Applicability and HOT Skills

	Equality and Fairness		Applicability		Evaluation HOT Skills	
Assessment	Accessible to all students	Requires Prior Knowledge	Time to Evaluate (minutes)	Materials Required	Spearman Coefficient	p-value
1	Yes	No	0.04	1/2 sheet of paper	0.39	0.06
2	Yes	No	0.94	1 sheet of paper	0.31	0.06
3	Yes	Yes	1.31	1 sheet of paper	0.35	0.17

Assessment 1 was the assessment that best met the criteria of being equally accessible to all students, requiring the least amount of materials and time to administer and evaluate, and finally having the highest correlation of scores to the performance-based assessment of the final project, an indication of the evaluation of HOT skills. The success of Assessment 1 in this evaluation may be highly dependent on the number and quality of the multiple-choice questions, however. The revisions to the assessment from

its original form resulted in scoring of the multiple-choice questions on a possible scoring

scale of 4-point rather than 12-points as on the Myers and Burgess (2003) assessment. As

a result, the ranking of the students' scores for Assessment 1 was crowded, causing many

tied scores. Additional multiple-choice questions should have been incorporated to

reduce the crowding in the rankings. The use of 12 multiple-choice questions would

increase the point scale to the same point scale used on the Meyers and Burgess (2003)

assessment. Additionally, the nature of multiple-choice questions focuses only on

whether or not the correct answer was selected and does not allow the student to explain

their reasoning for selecting the correct answer (Kohn, 2000). This allows for removing

ambiguity when grading, but it is at the expense of the extensive and descriptive

assessment data obtained from other forms of assessment (Harrison (2014). Both the

crowding of the student rankings and the reduction in the descriptive assessment data that

resulted from changes to the assessment should be considered in future use of

Assessment 1.

Assessment 2 also met the criterion of equality and fairness, required a modest

amount of materials and only had a slight decrease in the degree of correlation to that of

the Assessment 1 (See Table 18). Assessment 2 was altered by including more LOT

questions than in the Hofstein et al (2005) assessment. A possible effect of including the

LOT questions would be that the LOT questions encouraged the students to look closer at

the critical reading portion. Assuming that it is probable that students scoring higher on

the HOT questions would in turn correctly answer the LOT questions, this would result in

an increase in the composite score. This should not have changed the rankings within the assessment but only increased the point scale of the assessment. Because the Spearman correlation coefficient is based off student rankings the degree of correlation should not have been affected.

The free-response portion of Assessment 2 allowed for extensive and descriptive assessment data from students developing their own questions and explaining their reasoning for selecting a question for further investigation. The result of the free-response portion is that the length of time to evaluate student responses increased by an average of 54 seconds per student from that of Assessment 1, making it less applicable.

Assessment 3 fit the least into the criteria evaluated in this report. It requires prior knowledge acquired as part of the semester curriculum to complete. This switches the focus of this assessment from evaluating HOT skills to evaluating the content of the semester, which is not the focus of this study. The assessment was altered to include more free-response questions. As a result of this change the assessment required more time to grade affecting the applicability of Assessment 3. The other effect was that with more HOT free-response questions there was a larger point scale. This resulted in a greater distribution of point rankings, a reduction in the number of tied rankings so that the correction factor for tied rankings was not required when calculating the Spearman's correlation coefficient the degree of correlation should not have been affected.

Limitations and Areas for Future Research

There are some possible points of improvement that could be addressed with further research. There is still a lot of debate as to the validity of using multiple-choice questions to evaluate HOT skills (Moss, 1992; Harlen, 2000; Erdogan, Campbell, & Abd-Hamid, 2011; Puncochar & Klett, 2013; Harrison, 2014). Further studies could explore if the validity of the assessments was compromised by changing some of the free-response questions to multiple-choice questions.

Harrison (2014) identifies an advantage of performance-based assessments over single-day paper-and-pencil assessments (like the three assessments in this study) is that the student is not assessed once but continually as they work through the inquiry activity. The continuous assessment technique compensates for variations in student performance from day to day whereas the single day paper-and-pencil assessment may be influenced by other factors depending on the day the student participates in the assessment. In this study, the final project was a semester long group project. A semester long group project requires cooperation with peers and resilience. So the final score of the individual is influenced to some extent to the people they are partnered with. Further studies could investigate if the scores from the shorter inquiry projects' performance-based assessments (that were not used in the study) would be different from a semester long inquiry performance-based assessment.

Summary

In summary, Assessment 1 and Assessment 2 met the requirements of the study. Both have advantages and disadvantages, yet both are practical tools that could be used to evaluate the effectiveness of inquiry-based teaching techniques with respect to developing HOT skills in tenth grade biology students if they are carefully written and administered.

Implications

I have employed both Assessment 1 and Assessment 2 and find that I am using Assessment 2 more often. I prefer Assessment 2 because it gives more depth of understanding that is acquired from evaluating students' ability to generate questions in Assessment 2. I tend to agree with Kohn (2000) "No matter how clever or tricky the questions are, multiple-choice tests simply do not measure the same cognitive skills as are measured by similar problems in free-response form" (Kohn, 2000, p. 8). By evaluating the number and type of questions the students judge as valuable, I am able to acquire more information about their thought processes such as misconceptions than is available from the multiple-choice questions in Assessment 1. Another advantage that I have observed from implementing Assessment 2 is that the nature of the assessment does not feel like another assessment. The students' curiosity is sparked by a novel experience and when they are asked to generate questions that interest them. Additionally, even though grading free-response questions requires more time per paper, creating HOT multiple-choice questions for Assessment 1 is difficult and very time consuming.

58

ASSESSING ASSESSMENTS

I have created and used a post-assessment (See Appendix H) to accompany Assessment 2 for pre- and post-testing throughout the year. I believe that by employing Assessment 2 as a self-evaluation tool, I will be able to determine the most effective inquiry lessons to maximize my effectiveness as a secondary education teacher. These inquiry lessons will hopefully encourage scientific literacy so that the next generation can make informed decisions.

References

Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education, 25*(1), 31-36.

Barrow, L. H. (2006). A brief history of inquiry: From Dewey to Standards. *Journal of Science Teacher Education, 17*(3), 265-278.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain* (Vol. 19, p. 56). New York: David McKay Co Inc.

Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*(1), 37-50.

Bridgeman, B., & Rock, D. A. (1993). Relationships among multiple–choice and open–ended analytical questions. *Journal of Educational Measurement, 30*(4), 313-329.

Clark, D. (2010). *Bloom's taxonomy of learning domains: The three types of learning*. Retrieved from http://www.nwlink.com/~donclark/hrd/bloom.html

Copper, K. & Kazilek, C. J. (2016). *Seeing color: Ask a biologist*. Retrieved from http://askabiologist.asu.edu/explore/seeing-color

East Bay Educational Collaborative (2015). *Scientist notebook*. Lincoln Rhode Island. Retrieved from http://ebecri.org/custom/Lincoln.Notebook.Format;

Erdogan, I., Campbell, T., & Abd-Hamid, N. H. (2011). The student actions coding sheet (SACS): An instrument for illuminating the shifts toward student-centered science classrooms. *International Journal of Science Education, 33*(10), 1313-1336.

Grady, J. (2010). The inquiry matrix. *The Science Teacher, 77*(8), 32.

Harlen, W. (2000). *Assessment in the inquiry classroom. Foundations (Volume 2): Thoughts, views and strategies for the inquiry classroom*, Retrieved from http://www.exploratorium.edu/files/texnet/profdevelopment/media/nsf99148.pdf#page=87

Harrison, C. (2014). Assessment of inquiry skills in the SAILS project. *Science Education International, 25*(1), 112-122.

Hasse, S., Joachim, C., Bögeholz, S., & Hammann, M. (2014). Assessing teaching and assessment competences of biology teacher trainees: Lessons from item development. *International Journal of Education in Mathematics, Science and Technology, 2*(3).

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching, 42*(7), 791-806.

Hohenshell, L. M., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed method study. *International Journal of Science Education, 28*(2-3), 261-289.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools.* Portsmouth, NH: Heinemann.

Lukhele, R., Thissen, D. and Wainer, H. (1994), On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement, 31: 234–250.*

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62(3), 229-258.*

Myers, M. J., & Burgess, A. B. (2003). Inquiry-based laboratory course improves students' ability to design experiments and interpret data. *Advances in Physiology Education, 27(1), 26-33.*

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

National Research Council (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington DC: National Academy Press.

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington DC: National Academy Press.

Next Generation Science Standards Lead States. (2013). *Next generation science standards: For states, by states*. Washington DC: National Academy Press.

Pechenik, J. A. (1996). *A short guide to writing about biology, third edition*. New York, NY: Addison Wesley Logman, Inc.

Puncochar, J., & Klett, M. (2013). A model for outcomes assessment of undergraduate science knowledge and inquiry processes. *Research & Practice in Assessment, 8,* 42-54.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49(3), 363-393*.

Roseman, J. E., DeBoer, G., Glassman, S., Goldstein, B., Hanselmann, P., Hardcastle, J., ..., Koppal, M. (2016). *The atlas of scientific literacy project 2061*. American Association for the Advancement of Science (AAAS). Retrieved from http://www.project2061.org/publications/atlas/

Rutherford, F. J., & Ahlgren, A. (1990). *Science for all Americans.* Washington, Dc. American Association for the Advancement of Science.

Saido, G. M., Siraj, S., Nordin, B., Bakar, A., & Al Amedy, O. S. (2015). Higher order thinking skills among secondary school students in science learning. *Malaysian Online Journal of Educational Sciences, 3(3), 13-20.*

Sampson V., Enderle P., Gleim L., Grooms J., Hester M., Southerland S., & Wilson K. (2014). *Argument-driven inquiry in biology lab investigation for grade 9-12*. Arlington, Virginia: National Science Teacher Association Press.

Shields, M. (2006). *Biological inquires: Standards-based labs, assessments and discussion lessons*. San Francisco, CA: Josey-Bass.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Madison, Wisconsin: National Institute for Science Education.

Zion, M., and Mendelovici R. (2012). Moving from structured to open inquiry; challenges and limits. *Science Education International. 23, 4, 383-399.*

ASSESSING ASSESSMENTS

Appendix A: Guided Inquiry Lesson Plan Osmosis

Concept/Topic to Teach:
- Osmosis, Plasmolysis, Turgor Pressure, Plant Cells

Objectives (Shields, 2006, p. 3**)**
- Learners engage in scientifically oriented questions.
- Learners give priority to evidence in responding to questions.
- Learners formulate explanations from evidence.
- Learners connect explanations to scientific knowledge.
- Learners communicate and justify explications.

Specific Objectives:
o Students will implement safe procedures and practices when manipulating equipment, materials within the laboratory.
o Students will analyze data related to the fluid mosaic model of cell membranes and the impact of osmosis.
o Students will design and conduct an experiment.
o Students will analyze the results form their experiment.
o Students will publish the result of their experiment.

Step-by-step Procedure:

Advance Preparation   **11-04-15**
- Bell ringer on the board.
- Investigating Osmosis in Plant cells Lab (Sheilds, 2006, p. 88-92)
  ➢ Elodea plant leafs
  ➢ Blank Slides
  ➢ Cover Slips
  ➢ Water with dropper
  ➢ Microscopes (12)
  ➢ Analytical Balance

Subject: Biology

Grade: 10th

Time: 4 days

Teach 21 Standards:
SC.O.B.2.6 Analyze the chemistry and fluid mosaic model of the cell membrane as they relate to import and export of molecules necessary for life including: Osmosis, Diffusion, Active Transport, Passive Transport, Dialysis
SC.O.CB.1.1 Implement safe procedures and practices when manipulating equipment, materials, organisms, and models.
SC.O.CB.1.2 Formulate scientific explanations based on historical observations and experimental evidence, accounting for variability in experimental results.
SC.O.CB.1.4 design, conduct, evaluate and revise experiments (e.g., compose a question to be investigated, design a controlled investigation that produces numeric data, evaluate the data in the context of scientific laws and principles, construct a conclusion based on findings, propose revisions to investigations based on manipulation of variables and/or analysis of error, or communicate and defend the results and conclusions).

ASSESSING ASSESSMENTS

> ➢ Salt
> ➢ Erlenmeyer Flask
> ➢ Beaker
> ➢ Direction sheets (not written on)

Procedure
- *Bell ringer* (5 min):
- *Pre Lab Instruction (15 min):*
    - o Hand out Lab direction sheets (not to be written on). Have the students follow along while you demonstrate the appropriate procedure for the lab. Check for understanding and allow them to being the lab with their assigned partners.
- *Application of Knowledge/Lab (35 min):*Investigating Osmosis in Plant Cells Lab: Part I
    - o Students will demonstrate skill of making a wet mount slide of Elodea plant. They will draw the cells under an appropriate magnification and will label their diagram appropriately. Students will then preserve the plant leaf and remove most of the water from the wet mount slide. They will then add 10% salt solution to the leaf and prepare it as a wet mound slide. Then the students will draw the cells under the same magnification and label their diagram appropriately. Finally the students will compare and contrast their observations of the leaf with 100% water and that in 90% water.

Guided Inquiry   **11-05-15**
Advance Preparation
- Bell ringer on the board.
- Investigating Osmosis in Plant cells Lab (Part II) (Shields, 2006, p. 88-92)
- Copies of Experimental lab proposal WS (source unknown)

Procedure
- *Bell ringer* (5 min):
- *Guided Inquiry (45 min):*Investigating Osmosis in Plant Cells Lab: Part II
    - o Students are to work with their assigned lab partner to design an appropriate Lab proposal. Due at the end of the hour. Question: What salt water concentrations would be an isotonic solution? Or What is the concentration of water in the cell so that it will not lose or gain water in an isotonic solution?

Advance Preparation   **11-06-15**
- Bell ringer on the board.
- Investigating Osmosis in Plant cells Lab (Part II) (Shields, 2006, p. 88-92)
    - ➢ Elodea plant leafs

ASSESSING ASSESSMENTS

> Blank Slides
> Cover Slips
> Water with dropper
> Microscopes (12)
> Analytical Balance
> Salt
> Erlenmeyer Flask
> Beaker
> Direction sheets (not written on)
> Copies of Experimental lab proposal WS (source unknown)

Procedure
- *Bell ringer* (5 min):
- *Guided Inquiry (45 min):*Investigating Osmosis in Plant Cells Lab: Part III
    - o Students will conduct their designed experiment. And record their results.

Advance Preparation   **11-10-15**
- Bell ringer on the board.
- Investigating Osmosis in Plant cells Lab (Part II) (Shields, 2006, p. 88-92)
- Copies of Experimental lab proposal WS (source unknown)
- Directions for poster publication of results (Sampson et al., 2014, p. 7)

Procedure
- *Bell ringer* (5 min):
- *Guided Inquiry (45 min):*Investigating Osmosis in Plant Cells Lab: Part III
    - o Students are to publish their results of their experiment by making a scientific poster that illustrates their; Question, Clam, Results, and Explanation of results. ASSESSMENT

Reference

Sampson et al. (2014). Argument-Driven Inquiry in Biology Lab Investigations for Grades 9-12, National Science Teacher Association Press, Arlington Virginia Page 7

Shields, M. (2006). Biology Inquiries Standards-Based Labs, Assessments, and Discussion Lessons. Jossey-Bass Teacher. San Francisco California. Pages 88-92.

Appendix B: Guided Inquiry Lesson Plan DNA

Concept/Topic to Teach:
- The structure and function of DNA.

Objectives (Shields, 2006, p. 3**)**
- Learners engage in scientifically oriented questions.
- Learners give priority to evidence in responding to questions.
- Learners formulate explanations from evidence.
- Learners connect explanations to scientific knowledge.
- Learners communicate and justify explications.

Specific Objectives:
- o Students will implement safe procedures and practices when manipulating equipment, materials within the laboratory.
- o Students will analyze data related to DNA.
- o Students will design and conduct an experiment.
- o Students will analyze the results form their experiment.
- o Students will publish the result of their experiment.

Required Materials
- Textbook –Miller, K. R. & Levine, J. (2006). *Biology*, Boston Massachusetts. Pearson Prentice Hall. If the textbook is lost or damaged it will be your responsibility to replace the textbook
    - o Pg. 287-308
    - o Pg. 322-332

Subject: Biology
Grade: 10th
Time: 2.5 days
Teach 21 Standards: SC.O.B.2.12 Construct and use models of DNA and explain replication and mutations.
SC.O.B.2.13 Differentiate the structure and function of messenger, transfer, and ribosomal RNA in the process of transcription and translation.
SC.O.B.2.14 Research and debate the application of DNA technology in the context of social, ethical and political issues.

Step-by-step Procedure:

Advance Preparation   **02-03-16**
- Bell ringer on the board.
- DNA Extraction
  - ➢ Strawberry (Fresh and frozen)
  - ➢ Cheese Cloth (Coffee filter, cotton balls, nylon)
  - ➢ Funnel
  - ➢ Test tube or beaker

ASSESSING ASSESSMENTS

➢ Rubbing Alcohol (Freezer)
➢ Wooden stir sticks
➢ Balance
➢ Extraction fluid (100ml dish soap, 15 g salt, 1 liter water)
➢ Activity write-up

Procedure
- *Bell ringer* (5 min): Is DNA the same in any cell in the human body? Explain your answer.
- *Explore: Pre Activity Instruction (15 min):* Demonstrate the steps that the students will be performing for the DNA extraction.
- *Explore: Guided Inquiry (30 min):* Have the students work with a partner to design and experiment (Question: How can we improve in the amount of DNA that is extracted?). They are to make a hypothesis, chose an independent variable, list the dependent and control variables, list the materials needed, and design the procedure. Some suggestions on variables that they could change include: frozen strawberry, fresh strawberry, warm strawberry; different type of filter examples coffee filter, nylon, cotton balls, cheese cloth; increasing or decreasing the smashing time; extending or reducing the filtration time; changing the amount of extraction fluid used.

Advance Preparation   **02-04-16**
- Bell ringer on the board.
- DNA Extraction
  ➢ Strawberry (Fresh and frozen)
  ➢ Cheese Cloth (Coffee filter, cotton balls, nylon)
  ➢ Funnel
  ➢ Test tube or beaker
  ➢ Rubbing Alcohol (Freezer)
  ➢ Wooden stir sticks
  ➢ Balance
  ➢ Extraction fluid (100ml dish soap, 15 g salt, 1 liter water)
  ➢ Activity write-up

Procedure
- *Bell ringer* (5 min): If you wanted to examine a suspect's DNA what cells would you use and why?
- *Explore; Guided Inquiry (50 min):* Students will be conducting the experiment they designed the previous day. They need to follow their procedure. Note: To determine the amount of DNA extracted they need to weigh the wooden stirring rod before they collect the DNA on it and then after when it has the DNA on it.

Advance Preparation   **02-05-16**
- Bell ringer on the board.

ASSESSING ASSESSMENTS

- Materials for production of a poster
  - ➢ 8 x 14 (legal) size paper
  - ➢ Rulers
  - ➢ Markers or color pencils
  - ➢ Direction sheet

Procedure

- *Bell ringer (5 min):* Why did we extract DNA from a Strawberry and not something else?
- *Explore (20 min)*: Students are to publish their results of their experiment by making a scientific poster that illustrates their; Question, Clam, Results, and Explanation of results. ASSESSMENT

Reference

Sampson et al. (2014). Argument-Driven Inquiry in Biology Lab Investigations for Grades 9-12, National Science Teacher Association Press, Arlington Virginia.

ASSESSING ASSESSMENTS

Appendix C: Semester long Open Inquiry Lesson Plan (Forest Ecology)

Concept/Topic to Teach:

      Inquiry oriented laboratory activities where students receive direct exposure to the investigative nature of science are invaluable to the students' long term retention. The focus of this lesson plan is to present several inquiry oriented laboratory activities that emulate those performed in the field of biology.

Objectives (Shields, 2006, p. 3**)**

- Learners engage in scientifically oriented questions.
- Learners give priority to evidence in responding to questions.
- Learners formulate explanations from evidence.
- Learners connect explanations to scientific knowledge.
- Learners communicate and justify explications.

Specific Objectives:

- Students will gain an understanding of what science means.
- Students will develop observational skills.
- Students will understand the steps involved in the scientific method.
- Students will study the steps in designing a good experiment.
- Students will relate science-technology-social issues while using variety of sources to construct and defend their solutions.
- Students will explore real world applications for their knowledge of Biology.
- Students will develop scientific skills of measurement related to the study of Biological Sciences.
- Students will demonstrate the ability to use the inquiry process to solve problems.
- Students will investigate how to evaluate data.

Step-by-Step Procedure:

Subject: Biology
Grade: 10th
Time: 10 class periods.
NGSS: S.10.LS.15 Students will create or revise a simulation to test solution to mitigate adverse impacts of human activity on biodiversity. S.10.LS.24 Students will evaluate the evidence supporting claims that changes in environmental conditions may result in: 1) Increase in numbers of individuals of some species 2) the emergence of new species over time and 3) the extinction of other species. Teach 21: SC.S.B.1 Students will: Demonstrate an understanding of history and nature of science as a human endeavor encompassing the contributions of diverse cultures and scientists. Demonstrate the ability to use the inquiry process to solve problems. Relate science-technology-societal issues while using a variety of sources to construct and defend their solutions.

71

Part 2: Laboratory Notebook Techniques   **8-31-15**

Objective
- Students will develop scientific skills of measurement related to the study of Sciences.
- Student will communicate using appropriate methodology.

Advance Preparation
- Bell ringer on the board.
- Copies of How to take Field notes.
- Field notebook from actual field experience (Write in the Rain from 2012 Utah trip 2014 Engineering class and 2015 Moose Watch Class)
- Copies of Writing Lab Reports (Pechenik, 1996)

Procedure
- *Bell ringer (5 min):*
- *Engagement (30 min):* As a class read the Writing Lab Reports article. Discuss
- *Explore (10 min)* Hand out actual field notebooks for them the explore (this may be the first time they ever see Write in the Rain paper, so demonstrate how it can get wet and not get damaged.)
- *Explain (10 min)* Cover examples of good and bad Field Notes.

Part 3; Vegetation Studies   **10-23-15, 11-4-15**

Objective
- Students will develop scientific skills of measurement related to the study of Sciences.
- Student will communicate using appropriate methodology.
- Students will gain an understanding of what science means.
- Students will develop observational skills.
- Students will understand the steps involved in the scientific method.
- Students will study the steps in designing a good experiment.

Advance Preparation
- Bell ringer on the board.
- Notebook, pencil, compass.

Procedure
- *Bell ringer (5 min):*
- *Guided Practice (10 min):* Hand out compasses to the students. Each student should orient themselves within the room (North, South, East, West). They are to practice taking compass readings from various predetermined locations within the room. Once the group is demonstrating proficient compass operational skills they are to proceed outside.
- *Project Based Learning (50 min):* Evaluate the area for suitability and any safety hazards, divide the students into groups of three. The students are to run a 10

meters long line oriented North and South. The students will document by making a simple plan drawing of key features – any nearby buildings, large plants (trees and shrubs), favored paths across the area, slopes, and so on. Include information along this line about the number and type of vegetation present. Make a note of any clearly visible features in the vegetation, such as areas of flowering plants, worn grass or darker vegetation. Student will then repeat the process with the 10 meter orientation line positioned East and West.

Background Information

When conducting Biological research it is often important to make observations that can be repeated. The technique of documenting observations so that they can be interpreted and compared with future and past observations is an invaluable technique, which frequently requires practice. Therefore it is advantages to take vegetation observations within specified locations to hone these skills.

Objective
- Students will develop scientific skills of measurement related to the study of Sciences.
- Student will communicate using appropriate methodology.
- Students will gain an understanding of what science means.
- Students will develop observational skills.
- Students will understand the steps involved in the scientific method.
- Students will study the steps in designing a good experiment.

Advance Preparation
- Bell ringer on the board.
- Notebook, pencil, compass.

Procedure
- *Bell ringer (5 min):*
- *Inquiry; Question (10 min):* After the students have made an initial observation of their study area they need to begin brainstorming possible questions that have occurred to them. This brainstorming process may require many days of bell ringers. (Examples could relate to the type of vegetation in their area, the health of the vegetation, evidence they noticed of animal life in their area, etc.)
- *Inquiry; Gather Evidence (50 min):* After the students have developed some basic questions they are to research their question, to gather evidence about their questions.

Objective
- Students will develop scientific skills of measurement related to the study of Sciences.
- Student will communicate using appropriate methodology.
- Students will gain an understanding of what science means.
- Students will develop observational skills.

ASSESSING ASSESSMENTS

- Students will understand the steps involved in the scientific method.
- Students will study the steps in designing a good experiment.

Advance Preparation
- Bell ringer on the board.
- Notebook, pencil, compass.

Procedure
- *Bell ringer (5 min):*
- *Inquiry; Propose a possible Explication (20 min):*After the students have developed some basic questions they are to research their question, to gather evidence about their questions.
- *Inquiry; Gather Evidence (40 min):* Monthly the students are to proceed to their sample area and record the same type of data.
  - o The students are to run a 10 meters long line oriented North and South. The students will document by making a simple plan drawing of key features – any nearby buildings, large plants (trees and shrubs), favored paths across the area, slopes, and so on. Include information along this line about the number at type of vegetation present. Make a note of any clearly visible features in the vegetation, such as areas of flowering plants, worn grass or darker vegetation. Students will then repeat the process with the 10 meter orientation line positioned East and West. (Photographing is optional but not a substitute for accurate drawings)

Objective
- Students will develop scientific skills of measurement related to the study of Sciences.
- Student will communicate using appropriate methodology.
- Students will gain an understanding of what science means.
- Students will develop observational skills.
- Students will understand the steps involved in the scientific method.
- Students will study the steps in designing a good experiment

Advance Preparation
- Bell ringer on the board.
- Notebook, pencil, compass.

Procedure
- *Bell ringer (5 min):*
- *Inquiry; Publish (50 min):* After collecting data for several months the students are to analyze their data and publish it. They will be required to prepare a Laboratory Report, Present a 10 minute Presentation, and create a Poster explaining their findings. ASSESSMENT
- *Inquiry; Add to Explanation (10 min):* Finally the students will have to Re-examine their explanation after hearing/seeing their peers' experiments.

ASSESSING ASSESSMENTS

Background Information
   When conducting Biological research it is often important to make observations that can be repeated. The technique of documenting observations so that they can be interpreted and compared with future and past observations is an invaluable technique, which frequently requires practice. Therefore this activity where students take repeated vegetation survey observations within specified locations will develop such skills.
   Another advantage to this year long study is to have the students become familiar with the changing features of nature. Over the course of a school year vegetation experiences general increases and decreases in abundance. Most people fail to observe these changes by making the students concentrate on observing these variations they will become more aware of making general observations throughout their lives.
   Finally this is a true inquiry activity where the students after making initial observations of their sampling area will develop from beginning to end the entire inquiry process. This type of activity is frequently impossible to facilitate in a classroom environment unless it is spread over a significant time period. By taking a day out of every month the students will begin to understand the time required when actually performing Biological Scientific Research.

Assessment Based on Objectives
Students will be assessed on their final project results will be published and assessed for: How appropriately they were able to implement the inquiry process, operate scientific equipment to take accurate scientific measurements, and appropriately communicate their results via laboratory reports, presentations and posters.

References
East Bay Educational Collaborative; http://ebecri.org/custom/Lincoln.Notebook.Format
   Scientist Notebook; East Bay Educational Collaborative; Lincoln Rhode Island,
   September 1st 2015.
Pechenik, J. A. 1996; A Short Guide To Writing About Biology, Third Edition, Pg. 52-
   60, New York, NY.
Rodney D., Chan F., Tamir P., and Lenhardt C.; 2002 Science Educator's Guide to
   Laboratory Assessment; NSTApress. Arlington VA
Shields, M.; 2006; Biology Inquiries Standards-Based Labs, Assessments, and
   Discussion Lessons. Page 3. San Francisco, CA

Appendix D: Inquiry Project Scoring Rubric

Inquiry Participation Assessment Project Rubric

Group Work			
	Smashing/ Genius/Excellent	Polished/Pretty Darn Good	Intact/Pretty Good	Unfortunate
Participation	Always on task (10)	Mostly present, on task. (5)	Frequently on task. (3)	Frequently off task. (0)
Preparation	Well rehearsed, written material clear, is ready to work. (10)	Prepared, but still rough procrastinates. (5)	Somewhat prepared, or prepares inconsistently. Not well organized. (3)	Unprepared, under-rehearsed, sloppy work. (0)
Attitude	Cooperative, receptive, positive, dependable, enthusiastic, ready willing able. (10)	Helpful to others when asked, involved - not always contributing to energy of class, ready to work. (5)	Physically present, does bare minimum. (3)	Negative, disrespectful, closed off to people's needs, tunnel vision, coming to class and doing other things. (0)
Creativity	Willing to risk and fail, tries new ideas, imaginative, unique approach. (10)	Moments of good ideas- not fully thought out, willing to try- but inconsistent inspiration, relatively open-minded, plays it safe. (5)	Having ideas-but not following impulses, copying with variations, not able to explore material. (3)	Unwilling to experiment with new ideas, copycat, plagiarism (0)
Problem solving	Flexibility, work 'on their feet' follows through with ideas. (10)	Attempting to fix problem, not always succeeding, inconsistent follow through, not always able to be flexible or is sometimes frustrated by process. (5)	Identifies problem, but can't solve it, get stuck, little or not follow through. (3)	Isn't aware of the problem, gives up on process, shows frustration or anger. (0)

ASSESSING ASSESSMENTS

Note Taking	Notes indicate quality researched notes. (20)	Organized and good quantity of notes (15)	Some notes (10)	No notes (0)
The Resources Used	All sources sited (5)	Most sources sited (3)	Few sources sited (1)	No sources sited (0)
Inquiry				
Make Observations	Observation notes are accurate and concise. Including measurements, drawings, and numerous observations (20)	Observation notes include some measurements, some drawing and some observations. (15)	Observation notes are minimal. Not clear with few if any measurements, drawings, or observations (10)	No observation notes (0)
Define Questions	Well thought out question generated. A question that requires extended time, research or collaboration to solve. (10)	Question that have more than one answer. (5)	Question that require cognitive discussions made by students. (3)	Recall question. (1)
Quality of your research (what you found out)	Excellent information that answers the guiding question fully and in a logical order (35)	Good information that answers most of the guiding question (20)	Some information given (10)	Little information given is difficult to understand (0)
Communication (Discussion)	Gives and receives feedback in a balanced way, aware of how other are affected.	Sometimes contributes to discussion, comments show inconsistent understanding.	Confused, but tries to talk anyway, contributes to discussion occasionally.	Rarely contributes to discussion, no understanding.

| Conclusion, Claims and Evidence	Conclusion stated clearly, incorporated evidence in your own words. Each claim supported by accurate evidence. All claims related to underlying question. (20)	Conclusion stated clearly and in your own words with some evidence given to support claims. (15)	Conclusion given in your own words with little to no evidence given. Claim not well supported. (5)	Conclusion given with little to no evidence given. Conclusion not clear. (0)
Quality of your presentation (how well you shared what you found out)	Oral presentation articulated clearly in your own words, with correct posture, clean speaking (not reading) and good volume. Visual presentation gives information in you own words, is neat, legible attractive with correct spelling. (20)	Oral presentation mostly articulated clearly. Visual work mostly presented clearly. Multimedia week mostly presented clearly (15)	Some oral presentation articulated clearly. Some of visual work presented clearly. Some of multimedia work presented clearly. (5)	Oral presentation not presented clearly. Visual work not presented clearly. Multimedia work not presented clearly. (0)
Consider New Evidence	Excellent information relevant to the guiding question fully and in a logical order (10)	Good information that answers most of the guiding question (5)	Some information given (3)	Little information given is difficult to understand (0)
Add to Explanation	Conclusion stated clearly, incorporated additional evidence given in your own words. Each claim supported by accurate evidence. All claims related to underlying question. (10)	Conclusion stated clearly and in your own words with some additional evidence given to support claims. (5)	Conclusion given in your own words with little to no additional evidence given. Claim not well supported. (3)	Conclusion given with little to no evidence given. Conclusion not clear. (0)

Appendix E: Assessment # 1

Earthworm Population Study

For a long time it has been observed and accepted that Earthworms are beneficial to the soil environment. They have been noted to increase the amount of air and water that get into the soil by creating tunnels and burrows. They break down and redistribute organic matter, such plant matter, as they consume the organic matter and relocate it in their castings (poop). These castings are an excellent source of fertilizer for plants, for they contain many nutrients that can be absorbed by the plants because the Earthworm digested the organic matter. Otherwise, the plants would not be able to absorb these nutrients trapped in the organic matter.

Because of the many benefits that earthworms contribute to the soil a farmer wanted to know how plowing his field affected the earthworm population. So the farmer decided to take several samples of earthworms over the course of two years in two locations; an unplowed field and a nearby plowed field. So as to not disturb the soil the farmer prepared a solution of mustard powder and water and applied this solution to the top of the soil in one square meter areas. The mustard solution irritates the earthworms so that they come to the surface where they can be counted. The mustard solution does not cause any lasting harm to the earthworms so that the experiment does not affect the earthworm population in the sampling areas.

Below is a chart depicts the number of earthworms found in the soil over the course of this study.

Use the information in the graph to answer the following questions.

79
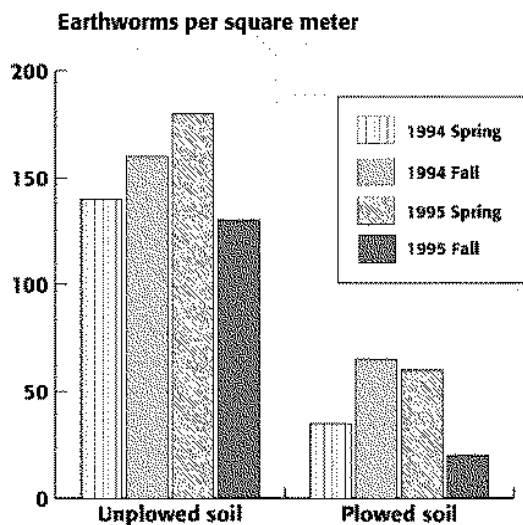
ASSESSING ASSESSMENTS

1. What question did the scientists who collected this data want to answer?
a. Are there more earthworms in the soil in the spring or in the fall?
b. What is the effect of plowing soil on the number of earthworms?
c. How is the size of earthworms affected by the seasons?
d. Does plowing soil affect how fast earthworms grow?

2. Where and when were the most earthworms found?
a. unplowed soil, spring 1995   c. unplowed soil, fall 1995
b. unplowed soil, fall 1994   d. plowed soil, spring 1994

3. What do the data in this graph show?
a. Unplowed soil has more earthworms than plowed soil.
b. Plowed soil has more earthworms than unplowed soil.
c. Plowing of soil has no effect on the number of earthworms.
d. The number of earthworms cannot be predicted.

4. A scientist who wants to study the effects of a new fertilizer on plants would probably
a. give each experimental group the same amount of the fertilizer.
b. not worry about measuring the amount of fertilizer used.
c. include a control group that received no fertilizer.
d. use different numbers of plants in each group.

Appendix F: Assessment #2 Part A: The Experiment

The Experiment: <u>Red Dot Special</u> (Shields, 2006)

Introduction
After staring at a red dot, if a person looks at white paper, he or she will see a greenish-blue dot that is not really there. This discrepant event captures students' attention, piques their curiosity, and launches them on an inquiry quest for an explanation. (Student will not have prior knowledge about the cause of such an event, making it ideal.)

What is Happening
The retina of the human eye consists of three types of color-sensitive neurons called cones. There are three different types of cone, and each responds best to a different part of the visible light spectrum. One type maximally absorbs light at 455 nm or blue light. Another best absorbs wavelengths at 530 nm or green light. The last type is most sensitive to red light at 625 nm. There is considerable overlap in the absorption ranges of the three. By comparing and analyzing impulses from the three types of cones, our brain is able to distinguish thousands of different colors.

Upon staring at a red dot for a prolonged period, a person exhausts the red-sensitive cones in a dot-shaped area of the retina. Then when looking at a white surface they see the bluish-green afterimage. The white light comes from the white surface contains all of the colors of the visible spectrum, including red. However, in the area of the retina that had been responding to a forming an image of the red dot, the exhausted red cones will not transmit and impulse to the brain. But the blue-sensitive and green-sensitive cones in the area will send signals to the brain. Thus, the brain will be tricked into perceiving a blue-green dot where one doesn't exist.

The red-sensitive cones become only temporary no responsive. Staring at a colored dot causes the cluster of cones to repeatedly fire, thus depleting their energy (ATP) supply. Energy is required by neurons to run the sodium potassium pump that "rests" the cell for a new action potential.
Material Per Student
- Index cards, white, unlined 4" X 6" or larger
- Adhesive red dots, 2 inch in diameter

Students Instruction
1. Begin by having the students experience the discrepant event. All at once have all students raise their cards so that they can see the red dot. Then have them quickly move the card so that the dot is about 20 cm from their eyes. They should stare at the dot for one minute (keep track of the time so that they don't have to). They should continuously stare at the dot while not moving the card and trying to minimize blinking.

2. At the end of one minute have the students flip their cards over and stare at one place on the pure white side. They should see a bluish-green dot.
3. Repeat the process one or two more times until most students have successfully seen the afterimage.

The Questionnaire (Hofstein et al., 2005)

Each student: Write your observations
After performing the experiment, answer the following questions:
1. What questions do you have after the experiment?
2. Choose one of those questions as an inquiry question.
3. Why did you choose that question?
4. Write a hypothesis that fits your inquiry question. The hypothesis is your expected answer to your inquiry question.
5. Suggest an experiment that can verify if your hypothesis is correct. In your suggestion justify the need for any stage of the experiment.

Analysis of Results based on Hofstein et al. (2005)
Low- and high-order-type questions related to the practical test and the critical reading of an article
Asking Questions   The Practical Test
        Critical Reading of an Article
Low-Order Questions   -What is A?
                                 -Which occurred?
                                 -Why did the color change?
                                 -What allows your eye to see in color?
High-Order Questions   -Is shade of the color influencing the results?
                                 -How does the amount of …?
                                 -What would occur if…?
                                 What is the relation between…?

Assessment #2 Part B: Critical Reading Article

By Kim Copper and CJ Kazilek (2016)
Illustrated by Dr. Biology

How we see color

You look out at a field of wildflowers showing off their bright reds, brilliant blues, and accents of yellow and white centers. These are just a few of the rainbow of colors you

will see today, but have you ever wondered how we see these colors? What about other animals, do they see the same colors as you? Do animals see color at all?

Field of colorful wildflowers displaying their bright red and blue colors. Image by Dellex via Wikimedia Commons.

You might not know it, but it is the light bouncing off objects like our field of flowers that gives us the ability to see. When the light enters our eyes, special cells tell our brains about the light. These cells are called photoreceptors. Light is made of little bits called photons. When the sun shines, trillions and trillions of these little bits of light fall on the earth. The photons bounce off of almost everything and some of them enter our eyes.

Those bits that enter our eyes allow us to see. So, where does the color come from?

Starting in the 1600s with Sir Isaac Newton, scientists have believed that there are different kinds of photons. Different types give rise to our sense of colors. The different photons are said to have different wavelengths. Sunlight contains all the different wavelengths of photons. The visible wavelength colors can be seen when you look at a rainbow. Raindrops acting as natural prisms bending the light to show the different wavelengths that produce the colors.

How do our photoreceptors work?

We have two main types of photoreceptors called rods and cones. They are called rods and cones because of their shapes.

83

ASSESSING ASSESSMENTS

These cells are located in a layer at the back of the eye called the retina. Rods are used to see in very dim light and only show the world to us in black and white.

This is why you see only black and white when you are outside in the evening or in a dimly lit room. The other type of photoreceptors, the cones, allow us to see colors. They are not as sensitive as the rods so they only work in bright light. There are three types of cones, one for each of the three main colors we see, red, green and blue.

Some people have a genetic defect that makes one or more of the cones fail. This condition is known as color deficiency. You may have heard it called color blindness. Color blindness is fairly common, affecting about nine percent of all humans. It is much more common in men than in women. To test for color blindness a special picture called an Ishihara test is used.

What about other animals? What kind of colors do they see? Most animals see fewer colors than we do, but some see more! We know this by looking at how many kinds of cone photoreceptors they have. Another good indication of what an animal can see is by looking at their own colors. The colors of their prey are also an indication of an animal's ability to see color.

Questions
1. What are the cells that receive light in our brain called?
    a. Light receptors
    b. Photoreceptors
    c. Photons
    d. Phenomenon
2. Having a genetic defect that prevents someone from having all three types of cone cells would result in.
    a. Complete blindness
    b. Color blindness
    c. Superior vision
    d. Not being able to see in dim light
3. Why do you see only in black and white when you are outside in the evening or in a dimly lit room?
    a. Because your eyes are tiered and do not work as well.
    b. Because rods are used to see in dim light and only show the world in black and white.
    c. Because the visible light is broken down by the raindrops to form a rainbow.
    d. Because the cones allow us to see color.
4. How can we determine what kind of color do animals see?

84

      a. Look at how many and what type of cone cells the animals has in their eye.
      b. Look at the color of the animal.
      c. Look at the color or the animals prey
      d. All of the above are indicators of the kind of color animals can see.

5. An animal that is active primarily at night such as a Lion would have more of what type of photoreceptor cells?
      a. Rods
      b. Cones
      c. Would not have either rod or cone cells.
      d. Would have an equal amount of rod and cone cells.

6. What type of photoreceptor cells were affected when a person looks at one type of colored dot for an extended period of time what part of the eye is affected?
      a. The rod cells are affected.
      b. The genetic make up of the eye is affected causing color blindness.
      c. The cone cells are affected.
      d. The photons are affected.

Appendix G: Assessment # 3

Part A: Multiple Choice
Identify the choice that best completes the statement or answers the question.

1. Why do scientists communicate their results in written reports?
    a. so other scientists will argue
    b. because they like to write
    c. to make scientific models
    d. so other scientists can repeat their experiments

2. Which is a step of inquiry method?
    a. analyzing results   c. using technology
    b. stating a theory   d. building a microscope

3. What is the only difference between the control group and the experimental groups in a controlled experiment?
    a. the test   c. the variable
    b. the prediction   d. the hypothesis

4. What is Ecology?
    a. The study of life.
    b. The study of populations.
    c. The study of living organisms in the natural environment, and how they interact.
    d. The study of energy and how it affects living things.

5. For the forest ecosystem, what is the main source of energy?
    a. The water   c. The soil
    b. The sun   d. The wind

6. For the forest ecosystem, what organisms can make their own food?
    a. Heterotrophs   c. All
    b. None   d. Autotrophs

7. For the forest ecosystem, what organisms do not make their own food?
    a. Heterotrophs   c. All
    b. None   d. Autotrophs

8. For the forest ecosystem, what animal eats both producers and consumers?
    a. Herbivore   c. Decomposer

    b. Carnivore   d. Omnivore

9. For the forest ecosystem, what animal breaks down dead decaying matter?
   a. Herbivore   c. Decomposer
   b. Carnivore   d. Omnivore

10. For the forest ecosystem, what the simplest group of organisms?
    a. Primary producer   c. Autotroph
    b. Primary consumer   d. Omnivore

Short Answer:
For the following list in complete sentences an example of the following that you observed in the forest ecosystem.
11. The ecosystem you observed; _____

_____

_____
12. A species you observed: _____

_____
13. A population you observed: _____

_____
14. A community you observed: _____

_____
15. A biotic factor you observed: _____

_____
16. An abiotic factor you observed: _____

_____
17. Describe a niche in the forest: _____

_____
18. What is a resource that there is competition for in the forest? _____

_____
19. Describe the movement of energy through the forest. _____

_____
20. Describe the food web in the forest. _____

_____

Part B: Survey Question:

Please answer the following using complete sentences. Take as much time as you need.

You have had two different forms of performing an experiment, one where you were given the question (How can we extract more DNA from a strawberry.) and one where you had to come up with your own question (The forest ecology project). Which did you like best? Why?

_____
_____
_____
_____
_____
_____

Appendix H: Post Assessment 2

Assessment #2 Part A: The Experiment

The Experiment: Linking Taste and Smell: Skittle Test

Introduction
Students working with a partner will close their eyes and plug their nose while they taste each flavor of skittles, recoding their observations in a data table. Students will then repeat the test but without having their nose plugged, recording their observations in the data table.

What is Happening
       Both your sense of smell and sense of taste detect chemicals. Your tongue is covered with about 10,000 taste buds, which detect five different kinds of tastes: salty, bitter, sweet, sour. You also have some taste buds on the roof of your mouth and inner surface of your cheeks.
       A person's sense of taste is greatly influence by the sense of smell. The chemical receptors involved in your sense of smell are located in a postage stamp-sized patch of nerve cells called the olfactory tract located at the roof of each nasal cavity. These receptors can detect up to a thousand different types of chemicals. Over 70% of what we think we taste actually comes from our sense of smell. When we eat, odor molecules travel between the mouth and the nose. The odor molecules meet with the olfactory receptor neurons in the nasal cavity and send a message to the brain.

Material per Pair of Students
-10 skittles (2 wild cherry, 2 strawberry, 2 raspberry, 2 berry punch, 2 melon berry)
-Data table to record observations with pencil

Students Instruction
1. Obtain a bowl of skittles. Choose one partner to do the tasting and the other partner to administer the test.
2. Have the partner taste each of the flavors with their nose plugged and eyes closed and ask them to guess the flavor.
3. Record their observations in the data table.
4. Then unplug their nose and have the individual smell each of the candies and guess the flavor while they taste. You may need to slightly smash the candy to increase the amount of scent produced.
5. Record their observations in the data table.

ASSESSING ASSESSMENTS

Actual Skittle Flavor	Subject's Guess with nose plugged	Subject's guess with nose unplugged

Questionnaire

Each student: Record your observations
After performing the experiment, answer the following questions:
6. What questions do you have after the experiment?
7. Choose one of those questions as an inquiry question.
8. Why did you choose that question?
9. Write a hypothesis that fits your inquiry question. The hypothesis is your expected answer to your inquiry question.
10. Suggest an experiment that can verify if your hypothesis is correct. In your suggestion justify the need for any stage of the experiment.

Analysis of Results based
Low- and high-order-type questions related to the practical test and the critical reading of an article and Asking Questions.

Low-Order Questions   -What is A?
                             -Which occurred?
                             -Why did the flavor change?
                             -What allows your tough to taste?
High-Order Questions   -Is eye sight influencing the results?
                             -How does the amount of …?
                             -What would occur if…?
                             -What is the relation between…?

Assessment #2 Part B: Critical Reading Article
What You See Is What You Taste, Says Scientist
Posted by Amy Briggs of National Geographic on April 12, 2013

When it comes to tasting, what you see is not always what you get. Speaking at the National Meeting of the American Chemical Society, Terry E. Acree, Ph.D., announced his findings that the appearance of foods and drinks can make people "see" flavors before they actually taste anything, a phenomenon that can influence their flavor experiences, food likes, and dislikes. Agee is hopeful that further understanding of how the eyes factor

into flavor perception can lead to the creation of healthy foods that will appeal to the pickiest of eaters.

The Eyes Have It

Traditionally, scientists have thought the tongue, nose, and brain dominated how people experience the flavors of the food, but Acree's work reveals how the visual can forcefully come into play. "Years ago, taste was a table with two legs—taste and odor," said Acree, who is with Cornell University's Department of Food Science.

"Now we are beginning to understand that flavor depends on parts of the brain that involve taste, odor, touch and vision. The sum total of these signals, plus our emotions and past experiences, result in perception of flavors, and determine whether we like or dislike specific foods."

Winning by a Nose

It's well known that smell can override a person's taste buds. Acree cited one popular experiment in which two groups of volunteers were asked to have a sip of plain water after smelling different foods. One group smelled sweet things like caramel and strawberries, while the other smelled non-sweet foods like bread, meat, or fish. For the sweet-smelling group, the plain water tasted sweet. But the water wasn't sweet at all for the other group. (Related: Secrets of Smell: Different Nose Parts for Stinky, Sweet.)

When a glass of white wine was tinted red, people's eyes changed how they tasted the wine. The natural chemicals that give merlots and cabernets their flavors came to the front. (Credit: André Karwath, Wikimedia Commons)

Your Lying Eyes

While the role of the eyes is important, it does not dominate all of a person's flavor perceptions. Acree pointed out that in different circumstances, other senses and parts of the brain can trump visual stimuli. For instance, certain foods—like hashes, chilies, and stews—can look like "vomit or feces" said Acree. But lots of people still eat (and enjoy!) these kinds of foods despite their unappetizing looks. So something beside the eyes is at work.

Acree posits a few different explanations. One is a person's memory: If a person has strong, positive feelings associated with these foods, that pleasant past experience can trump the yucky-looking visual on the plate. Another explanation Acree puts forward is

ASSESSING ASSESSMENTS

ASSESSING ASSESSMENTS

that people have a strong desire for new experiences and that input from the brain and nose may override the eyes in these situations.

By getting to the heart of how people's eyes, nose, brain, and tongue interact when eating, Acree believes that eventually we'll be able to develop healthy foods that are more appealing to a broader range of people—especially kids and picky adults.

Questions
1) Agee is hopeful that further understanding of how the eyes factor into flavor perception can lead to _____.
   a) a noble peace prize.
   b) the creation of healthy foods for the pickiest eaters.
   c) the creation of fast food for the pickiest eaters.
   d) continued employment for his research.
2) Flavor depends on the parts of the brain that also deals with
   a) Taste
   b) Odor
   c) Touch
   d) Vision
   e) All of the above
3) When the study group smelled something sweet before drinking water what did they taste?
   a) Vomit
   b) Red
   c) Sweet
   d) Water
4) According to Acree, while the role of the eyes is important, it does not dominate all of a person's flavor perceptions. What else affects a person's flavor perception?
   a) Past experiences.
   b) Its temperature.
   c) Its packaging.
   d) Its price.
5) What is most likely to occur if someone ate a food that looks very appetizing?
   a) The person would love the food.
   b) The person will not eat much of the food and be healthy.
   c) The person is more likely to eat the food.
   d) The person would never eat the food again.
6) According to Acree's research what would be the result if skittles were all the same color or not colored at all?
   a) Nothing, the skittles would be the same. Coloring does not have a taste.
   b) The skittles would not taste as good.

c) The skittles would taste better. People like to be surprised.
d) The skittles would smell different.