Brigham Young University
BYU ScholarsArchive

Theses and Dissertations

2010-05-03

Assessing Scientific Inquiry: Teacher Beliefs and Practices

Adam James Mitchell
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Biology Commons

BYU ScholarsArchive Citation
Mitchell, Adam James, "Assessing Scientific Inquiry: Teacher Beliefs and Practices" (2010). *Theses and Dissertations*. 2521.
https://scholarsarchive.byu.edu/etd/2521

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Assessing Scientific Inquiry: Teacher Beliefs and Practices

Adam J. Mitchell

A thesis submitted to the faculty of
Brigham Young University
In partial fulfillment of the requirements for the degree of

Masters of Science

Nikki L. Hanegan, Chair
Scott L. Howell
C. Riley Nelson

Department of Biology

Brigham Young University

August 2010

Copyright © 2010 Adam J. Mitchell

All Rights Reserved

ABSTRACT

Assessing Scientific Inquiry: Teacher Beliefs and Practices

Adam J. Mitchell

Department of Biology

Masters of Science

Science education reform movements have long urged the use of inquiry methods in all science instruction. More recently, standards and accountability reform efforts have emphasized measuring and improving student science achievement. Researchers have questioned the alignment and balance between these reforms (Lane, 2004; Yeh, 2001). This study addresses issues faced by secondary science teachers as they simultaneously meet the goals of these reform movements. Mixed methods were used to answer the questions: 1) Can a teacher's beliefs and practices regarding inquiry teaching methods be correlated with his/her assessment practices?; 2) What item types are most commonly employed by teachers that use an inquiry pedagogy?; and 3) What assessment strategies do teachers describe to assess scientific inquiry? Secondary science teachers, mostly from one western state, responded to a survey (N = 83) and provided a teacher-made classroom assessment (n = 30). Survey responses were used to assign a teacher inquiry score based on described frequency of pedagogical practices supporting or detracting from an inquiry focus. A rubric based on cognitive complexity was used to determine a numeric value for each test item with the sum of item scores providing an overall assessment score. Using regression analysis and Pearson's correlation this study found a moderate correlation ($r = 0.0447$, $p = 0.0133$) between teacher inquiry scores and assessment scores. A modest correlation was also established between teacher inquiry levels (high, medium, and low categories assigned using cut scores) and overall assessment scores using an ANOVA (DF=2, $p = 0.0262$) and Tukey-Kramer pairwise analysis (low to medium $p = 0.046$; low to high $p = 0.057$). Correlations indicate that teachers are able to simultaneously focus on inquiry in pedagogical and assessment practices. Cognitively complex items used by teachers with an inquiry focus measure the same cognitive skills as scientific inquiry. Survey responses to open-ended questions provided additional qualitative data supporting the study's findings. Respondents reported challenges in creating assessments that measure student scientific inquiry competency, but also noted that labs, observation and questioning, and performance assessments are useful in measuring inquiry skills.

Keywords: assessment, cognitive complexity, inquiry, secondary science teachers, science education reform

ACKNOWLEDGMENTS

I acknowledge my wife, Shannon M. Mitchell, for her love, patience, and constant support throughout my educational endeavors. My children, Elizabeth, McKenzie, and Xavier, have also provided considerable encouragement and endured well the incalculable time that I have spent away from them as I worked on this project. I would also like to thank the members of my graduate committee for providing substantial assistance and showing great kindness to me throughout this project. Dr. Nikki L. Hanegan for encouraging me to enter graduate school and for seeing me through to the end; Dr. C. Riley Nelson and Dr. Scott L. Howell for insightful feedback throughout each stage of this study. Christian Davies, David Kent, Shalene Hardman, Amber Bigler, and Sarah Rogers each spent time helping revise my writing and process data for this study. I would also like to thank the teachers who provided the data for this research.

TABLE OF CONTENTS

Abstract ii
Acknowledgements iii
Table of Contents iv
List of Tables v
List of Figures vi
Introduction 1
Literature Review 2
Scientific Inquiry 2
Science Assessment 6
Research Questions 10
Methods 10
Overview 10
Respondents 11
Survey 12
Classroom Assessment Documents 14
Analysis and Findings 21
Teacher Inquiry Levels, Assessment Scores, and Presence or Absence of Item Formats 21
Correlation between Teacher Inquiry and Assessment Scores 24
Teacher Responses to Open-ended Survey Questions 26
Discussion 36
Correlation between Teacher Inquiry Levels and Classroom Assessment Practices 37
Item Types used by Teachers with an Inquiry Focus 38
Assessment Methods Teachers Use to Assess Student Competency in Scientific Inquiry 39
Limitations 41
Future Research 41
References 43
Appendix A – Teacher Survey 49
Appendix B – Sample Items from Assessments 61

LIST OF TABLES

Table 1 *Teacher inquiry levels, assessment scores, and presence or absence of item types* 23
Table 2 *Results of the Tukey-Kramer pairwise analysis.* 26
Table 3 *Question 13 response categories.* 28
Table 4 *Question 14 response categories.* 29
Table 5 *Question 15 response categories.* 30
Table 6 *Question 17 response categories* 33
Table 7 *Question 18 response categories.* 34
Table 8 *Question 20 response categories.* 36

LIST OF FIGURES

Figure 1. Item scoring rubric 16
Figure 2. Correlation between test item score and teacher inquiry level. 25

Introduction

Science education reform movements have long focused on improving science curriculum, student learning, and the teaching of science through implementation of inquiry teaching methods and classroom activities that provide opportunities for students to engage in scientific inquiry (Anderson, 2007). More recently, the standards and accountability reform movement has called on science educators to expand their focus on measuring and improving student achievement in science (Britton & Schneider, 2007; Darling-Hammond, 2004). Bell (2007) indicated that assessment has a powerful influence on curriculum, pedagogy, and student learning. However, many researchers have questioned alignment between the push to use inquiry methods in teaching and a focus on assessment and accountability (Lane, 2004; Yeh, 2001). Orpwood (2001) points out that it took twenty years from the time that scientific inquiry skills were emphasized in national standards and curriculum goals until performance assessments were developed for use on large-scale science assessments. Wideen, O'Shea, Pye, and Ivany (1997) documented how implementation of science accountability assessments moved teachers away from inquiry activities and led to increased lectures, memorization, and test preparation.

Researchers (Aydeniz, 2007; Parx et al. 2004; Duschl & Gitomer, 1997) have explained that educational reform is a complex process with many obstacles, both structural and professional, limiting successful implementation. The dual emphasis on inquiry and assessment in science education reform increases the challenges of successful implementation of either reform movement (Trester & Jones, 2003). While education researchers often focus primarily on one of these goals, science teachers are constantly pressured to make simultaneous improvements in the use of inquiry pedagogy and assessment practices.

This study evaluated the connection between secondary science teachers' use of inquiry teaching methods and classroom assessment. Teacher-made classroom assessments were

examined for correlation with teachers' inquiry beliefs and instructional methods. The result is a description of teachers' beliefs concerning effective assessment of inquiry-based instruction. Additionally, this study presents data on the actual assessment practices of secondary science teachers and describes tools they are using to assess students' ability to perform scientific inquiry.

Literature Review

Scientific Inquiry

In science education reform few movements have been as long-lived or widespread as the push for inquiry teaching and learning. Throughout the last 50 years inquiry has served as a major theme in efforts to improve science curriculum, learning, and teaching (Anderson, 2007). A continued focus on inquiry is warranted since the primary objective of inquiry activities is to provide students with opportunities to learn scientific reasoning skills (Chinn & Malhotra, 2001). Project 2061 from the American Association for the Advancement of Science (AAAS, 2006) developed *Benchmarks for Science Literacy* (1993), standards describing the scientific knowledge and skills students should acquire before completing grades 2, 5, 8, and 12. These standards emphasize that students need to be able to explain how inquiry allows science, a distinct method for gaining knowledge about the natural world, to progress. The National Research Council's (NRC) *National Science Education Standards* (1996) and *Inquiry and the National Education Standards* (2000) both encourage science educators to provide opportunities for students to conduct their own scientific inquiry. Science education reform efforts led by the AAAS and the NRC have implied that inquiry practices of professional scientists serve as standards to guide teachers as they help students become scientifically literate adults (Champagne, Kouba, & Hurley, 2000). Because state agencies have relied heavily on the

Benchmarks for Science Literacy and *National Science Education Standards* for science curriculum standards, they have regularly included a significant emphasis on scientific inquiry in their own curriculum standards (AAAS, 2006).

Science educators and researchers have employed the concept of inquiry to identify a variety of educational practices. The term inquiry is often applied in one of three main ways (Bybee, 2000). Inquiry is used to describe: an element of scientific investigation, a process of learning, and a set of teaching methods. The term scientific inquiry routinely represents "the work of scientists, the nature of their investigations, and the abilities and understandings required to do this work" (Anderson, 2007, p. 808). In this sense scientific inquiry is foundational to the epistemology of science. Inquiry learning has been characterized as constructivist learning - a process in which individuals actively construct their own meaning for new ideas and concepts based on their previous understanding, the current context of the learning taking place, and social interactions with others. While it takes many forms, inquiry teaching is exemplified by teachers providing opportunities for students to generate authentic scientific questions and seek knowledge and understanding through hands-on problem solving activities. Because the goal of this study is to describe teacher beliefs and practices, we will focus on inquiry as an essential strategy of effective science teaching.

Chinn and Malhotra (2001) argue that, although providing opportunities for students to hone their scientific reasoning skills is one of the major goals of science education, most inquiry activities that students participate in at school do not accurately model the cognitive complexity of authentic science. Chinn and Malhotra present a framework to evaluate the authenticity of inquiry activities. Like other science education researchers, (see for example Bell, Smetana, & Binns, 2005; Hanegan & Friden, 2009; Windschitl, 2003) Chinn and Malhotra describe an

authenticity continuum on which inquiry tasks are placed - with authentic scientific inquiry at one end and simple inquiry tasks at the other. Here authentic scientific inquiry is used to describe research typical of professional scientists. Although authentic scientific research may make use of a variety of research methods and tools, in general it is "a complex activity, employing expensive equipment, elaborate procedures and theories, highly specialized expertise, and advanced techniques for data analysis and modeling" (Chinn & Malhotra, 2001, p. 177). On the other hand, simple inquiry activities, including simple experiments, simple observations, and simple illustrations, all lack the complex cognitive processes which are hallmarks of authentic scientific research.

Researchers have routinely sought ways to classify classroom inquiry activities (Bell et al., 2005; Chinn and Malhotra, 2001; Herron, 1971; Germann et al., 1996; NRC, 2000; Schwab, 1962; Tafoya et al., 1980; Windschitl, 2003). Four distinct levels of inquiry: confirmation experiences, structured inquiry, guided inquiry, and open inquiry have routinely been used to categorize classroom inquiry activities.

Confirmation experiences, referred to as "simple illustrations" by Chinn and Malhotra (2001), are defined as the least authentic form of inquiry activity. In these activities students verify a stated scientific principle by following a set of cookbook-like instructions. Although confirmation experiences provide important opportunities for manipulation of scientific equipment, these hands-on activities do not give students opportunities to generate questions, modify procedures, select variables to control and measure, or explain results.

In structured inquiry students are provided with a question, to which they do not know the answer, and a set of procedures designed to help them answer that question (Windschitl, 2003). Others have referred to this form of scientific inquiry as a "simple experiment" (Chinn &

Malhotra, 2001) and described that such an experiment typically investigates how the manipulation of a single independent variable influences a dependent variable. During structured inquiry activities, students also gain responsibility for explaining the experimental results.

Guided inquiry provides students additional opportunities for autonomy throughout the scientific process. For example, students are given a specific problem or question to investigate as well as access to a limited set of data collection equipment, but have the freedom to design their own experimental procedures, select their own variables, and explain the results (Windschitl, 2003).

Open or authentic scientific inquiry most closely matches the work of scientists. While limitations in resources such as time, equipment, money, space, and expertise require that teachers develop tasks that are simpler than those engaged in by professional scientists, the goal of authentic inquiry is to include as many of the essential elements of scientific inquiry as possible (Chinn and Malhotra, 2001). Therefore, in open inquiry investigations students formulate questions that are of interest to them from within a broad area of subject matter identified by the teacher. Students then design experimental procedures, including choosing which variables to control, manipulate, and measure, and collect appropriate data. Importantly, students must also interpret data collected within the framework of existing scientific theories (Windschitl, 2003).

Teachers face many challenges in attempting to implement authentic inquiry activities in their teaching (Anderson, 2007). However, Bol and Strage (1996) indicate that there is a general consensus among science education researchers, policymakers, and science teachers that inquiry-based teaching methods have the greatest ability to prepare today's students to be scientifically

literate at a time when scientific knowledge and associated technologies are advancing at an astonishing rate.

Science Assessment

In recent years, throughout local, national, and international educational systems, there has been an enlarged focus on science assessment (Britton & Schneider, 2007; Johnson & Hanegan, 2006). A significant goal of the No Child Left Behind Act (NCLB) of 2001 is improving student achievement through standards and accountability (2002). NCLB legislation required states to develop science content standards by the 2005-06 school year and to report student science progress using tests aligned to those standards beginning two years later (Le Floch et al., 2007). While a focus on science standards and assessment have increased practitioner and public awareness of the need for quality science instruction (Britton & Schneider, 2007), researchers are not convinced that these efforts are aligned with reform focused on increasing inquiry pedagogy (Anderson, 2007; Lane, 2004; Yeh, 2001).

All states have made efforts to meet assessment requirements established by the NCLB legislation (Le Floch et al., 2007). However, some science educators have expressed concern that these large-scale science tests focus on meeting the requirements of NCLB to the exclusion of measuring students' scientific reasoning skills (Lane, 2004). Historically large-scale assessments have employed multiple-choice and short answer items extensively. Although roughly half of states currently include at least some form of constructive response items, many states are also utilizing a considerable number of multiple-choice items on their standardized science assessments. These item types are capable of assessing a wide variety of content areas in a short amount of time and are easily scored (Miller, Linn, & Gronlund, 2009, p. 202-204). However,

researchers have argued that additional item formats are needed to accurately assess complex reasoning skills, needed to carry out scientific inquiry, encouraged by state and national science standards (Britton & Schneider, 2007; Lane, 2004; Songer & Gotwals, 2004; Yeh, 2001).

While standardized science tests are beyond the scope of this study, many researchers have noted a link between large-scale accountability assessments and classroom teaching and assessment methods. Bell (2007) described the influence assessment has on curriculum, pedagogy, and student learning. Following implementation of standardized assessments for accountability purposes teachers regularly modify their instruction to focus on items that will be on the test rather than the corresponding standards (Enger, 1997; Shavelson, Baxter, & Pine, 1991).

Researchers documented that following implementation of a high-stakes science assessment teachers moved away from inquiry activities and devoted more instructional time to lectures, memorization, and practice tests (Wideen et al., 1997). However, a survey of 257 teachers directly impacted by the introduction of a state-mandated high-stakes performance assessment identified several benefits from using this test format. Vogler (2002) found that these teachers made significant pedagogical changes to help students develop higher-level thinking skills in response to implementation of a performance assessment. Changes included teachers increasing the use of open-ended questions, posing questions that required critical thinking, emphasizing problem-solving activities, and allowing students to participate more frequently in inquiry investigations.

Shavelson, Baxter, and Pine (1991) outlined three forces: a shift towards a constructivist theory of learning, the push to include more hands-on learning experiences in science curriculum, and a general feeling that multiple-choice items are limited in what they can

measure, that are encouraging the inclusion of alternative item types on science assessments. Performance assessments have been shown to measure science constructs which are distinct from those measured by traditional item types (Shavelson, Baxter, & Pine, 1991). Likewise, laboratory tests and investigations measure knowledge that is considerably different than constructs measured by either multiple-choice or open-ended questions (Lawrenz, Huffman, & Welch, 2001). Others have argued that even if using a variety of item formats does not guarantee multiple constructs will be measured, it helps ensure that assessments are a fair measure of student understanding (Hollingworth, Beard, & Proctor, 2007).

The palette of item formats probing science problem-solving skills rather than factual recall is continually expanding. Zachos, Hick, Doane, and Sargent (2000) identified the "absence of objective assessment" as one of the main criticisms for programs attempting to develop student understanding of scientific inquiry. They created an assessment instrument designed to measure student ability to conduct scientific inquiry. Songer and Gotwals (2004) also asserted the need for assessment instruments capable of measuring complex scientific reasoning skills. They argued that few tests accurately assess the complex reasoning demanded by scientific inquiry. They developed an assessment, consisting of multiple-choice and open-ended tasks, which included items that had been mapped using a "content-inquiry matrix" to cover a broad range of content complexity and inquiry skills. Using this assessment they provided a more detailed analysis of students' changing science knowledge and inquiry skills.

Liu, Lee, Hofstetter, and Lin (2008) noted that past efforts to measure complex scientific thinking have been frustrated by: dichotomous scoring, scoring items for quantity of response rather than quality, and creating items that measure logic rather than reasoning. With these problems in mind they created "knowledge integration assessments" which include both

multiple-choice and constructed response items. Additional sensitivity was included in scoring rubrics to more accurately differentiate levels of student understanding on open-ended items. They concluded that this assessment system more accurately measured students' scientific knowledge and performance than traditional item types.

Although newly designed assessments are allowing researchers to measure student understanding of science inquiry, most assessment of science learning is still performed by the classroom teacher (Bell, 2007). Therefore, it is essential to understand how practitioner inquiry and assessment beliefs impact classroom practices. Using observation and interviews, Lorsbach, Tobin, Briscoe, and LaMaster (1992) described how two teachers' inquiry teaching methods related to their assessment practices. They found that inquiry teaching did not always lead to inquiry in assessment. One teacher provided many opportunities for students to use inquiry methods during laboratory assignments. However, his assessments focused on understanding science facts. A second teacher employed non-traditional assessment methods as she transitioned to using more inquiry in her teaching. Open-ended concept maps gave her students freedom to express understanding. Small-group oral examinations provided additional opportunity for students to show what they had learned. Oral exams allowed the teacher and students to arrive at consensus about the intended meaning of words used to describe science concepts.

Bol and Strage (1996) identified a significant gap between teachers' instructional goals and assessment practices. They interviewed 10 high school biology teachers to determine their assessment philosophies and practices. They also collected and analyzed classroom assignments and assessments to measure alignment between instructional goals and teacher practices. While science teachers routinely adopt instructional goals that focus on critical thinking and science inquiry, teachers rarely assess such capabilities. Instead teacher-made classroom assessments

tended to reinforce low level thinking and factual recall. Teachers were largely unaware of the gap between their instructional goals and assessment practices.

Science curriculum standards emphasize the need to employ inquiry teaching methods. Yet standardized accountability assessments focus on conceptual knowledge rather than inquiry skills. Although researchers have identified the need to improve the assessment of scientific inquiry, many questions remain. Science education researchers must better understand: how science educators view the gap between a focus on teaching scientific inquiry and the lack of emphasis on assessing inquiry, what methods classroom teachers use to assess inquiry, what impact large-scale science assessments have on teacher assessment beliefs and practices, and what relationship exists between teachers' use of inquiry teaching methods and the item formats included on classroom assessments.

Research Questions

Specifically, this study seeks to address the following questions:

1. Can a teacher's beliefs and practices regarding inquiry teaching methods be correlated with his/her assessment practices?
2. What item types are most commonly employed by teachers that use an inquiry pedagogy?
3. What assessment strategies do teachers describe to assess scientific inquiry?

Methods

Overview

This study employed mixed methods to investigate teachers' beliefs and practices regarding scientific inquiry and assessment. The questions of this study are complex and dynamic, as is

common in many areas of educational research, and were most appropriately addressed using the interdisciplinary and complementary methods afforded through mixed research (Johnson & Onwuegbuzie, 2004). To effectively address both quantitative and qualitative aspects of the research questions, this study gathered empirical evidence from a teacher survey and example classroom assessments.

Respondents

A total of 602 secondary science teachers (grades 7-12) were solicited for participation in this study using an e-mail invitation with a link to the survey. One hundred and one teachers responded to the invitation to complete the research survey (response rate of 16.8%), but only the 83 teachers that completed all sections of the survey (completion rate of 13.8%) were included in this study. No compensation was provided to survey respondents. Most teacher respondents (96.4%) were from a western state. One respondent was from a neighboring western state and two respondents were from states in the southern United States. These results mirrored the fact that a majority (97.2%) of available e-mail addresses were for teachers from one western state and a minority of addresses were from teachers in four other states. Female teachers represented almost half (45.8%) and male teachers slightly more than half (54.2%) of the sample.

Nearly all teachers (94%) identified themselves as “white.” A few teachers also identified themselves as “Pacific Islander” (2.5%), “Asian” (2.5%), and “black” (1%). All respondents were certified to teach science and the majority (65%) reported having an advanced degree in science or science education. Respondents’ teaching assignments at the time of this survey included biology/life sciences (64%), Earth science/geology (58%), chemistry (23%), and physics (12%). (Note: the previous percentages sum to greater than 100% as survey instructions

asked respondents to “mark all that apply” and many respondents were teaching multiple science content areas.) Respondents reported a wide range in teaching experience from 2 to 34 years (mean = 14.9 years, SD = 7.62). The pool of possible respondents was large and diverse, but the fairly low response rate introduces the possibility of sampling bias.

Responding teachers represented schools in urban (25.3%), suburban (49.4%), and rural communities (25.3%). Schools served by these teachers ranged in size from 160 to 3050 students (mean = 1108 students, SD = 632). The socioeconomic status of each school’s student population was determined by the percentage of students qualifying for free or reduced lunch. Teachers reported that a large majority of the schools where they taught had less than half of the student population qualifying for free or reduced lunch (<25% = 25 schools, 26 – 50% = 36 schools). Fewer respondents indicated that over half of their student population qualify for free or reduced lunch programs (51- 75% = 12 schools, 76 – 100% = 5 schools). Five respondents did not report the socioeconomic status of their students. The combined racial composition for students taught by survey respondents was reported as “white” (70%), “Hispanic” (17%), “American Indian” (4%), “Pacific Islander” (4%), “Asian” (3%), and “black” (2%). According to statistics available from the Department of Education from the western state where most survey respondents taught, the racial composition of students enrolled in public schools in this state during the year this survey was completed were “white” (79%), “Hispanic” (14.5%), “American Indian” (1%), “Pacific Islander” (1.5%), “Asian” (2%), “black” (1.5%), and “unknown” (0.5%).

Survey

The survey instrument used in this research included three main sections (see Appendix A). In Part I teachers were asked to describe individual and school characteristics. During Part II

teachers rated their scientific inquiry and assessment beliefs using Likert items. Open-ended questions in Part II allowed teachers to further explain their beliefs regarding the use of inquiry teaching methods and the assessment of scientific inquiry. Part III asked respondents to categorize the frequency of several common pedagogical and assessment practices on a Likert scale and provide additional details on these topics in response to open-ended questions. All demographic and Likert items included in the survey instrument were obtained from the National Survey of Science and Mathematics Education Science Questionnaire (Horizon Research, Inc., 2000). Open-ended items were developed by the researchers to provide additional qualitative data from which to draw conclusions. The survey addressed the topics of scientific inquiry and assessment making it possible that respondents held stronger opinions about these topics than a random sample of science teachers.

A total of 53 Likert items were included in Parts II and III of the survey. A simple coding scheme of “generally supportive of scientific inquiry,” “generally not supportive of scientific inquiry” and “not applicable” was developed based on prior classification schemes used to define levels of scientific inquiry in classroom activities (Bell , Smetana, & Binns, 2005; Chinn and Malhotra, 2001; Herron, 1971; Germann, Haskins, & Auls, 1996; NRC, 2000; Schwab, 1962; Tafoya, Sunal, & Knecht, 1980; Windschitl, 2003) as well as researcher experience from work on an earlier study (Morrison, 2008).

After developing the coding scheme, a panel of four science education researchers coded each of the Likert items. Each rater independently coded each item. Item codes were accepted when three or more raters agreed on the rating. Disagreements between the raters were resolved through group discussion until consensus was reached. Using this method, twenty-three of the Likert items were determined to be supportive of inquiry teaching methods, 22 were found to

detract from an inquiry focus, and eight items were seen as neither clearly supporting nor detracting from inquiry methods. The eight ambiguous items were not considered in further analysis.

Using responses to Likert items from Parts II and III of the survey, teachers were assigned a global inquiry level score. Points were awarded such that teachers who described frequently employing inquiry methods in their teaching received the highest scores. Cut scores were then employed to assigned teachers to one of three inquiry activity levels: high, medium, or low.

Content analysis was used to systematically categorize responses to the six open-ended questions included in Parts II and III of the survey (Stemler, 2001). Emergent coding was used for five of the six items. An initial review of all responses identified common themes and created coding categories that encompassed all responses. These categories were then used to code the original teacher responses using NVIVO, a qualitative analysis software package. One open-ended item, which asked teachers to "give an example of a typical inquiry activity" used in their classroom, was coded using a priori codes based on the levels of inquiry activities described by Chinn and Malhotra (2001) and Windschitl (2003).

Classroom Assessment Documents

During the second phase of this study, all 83 survey respondents were invited to submit an example of a classroom assessment used during the year they completed the survey and which exemplified "normal assessment practices" in their classroom. Thirty teachers submitted a total of 37 assessment documents including unit tests and quizzes (28), laboratory experiments (3), and rubrics for performance assessments (6). Most teachers (93.3%) provided an electronic version of their assessment. Two teachers submitted paper versions of their assessments.

Assessments had between 1 and 53 items (mean = 19.6, SD = 13.7) and represented the following science curriculum areas: chemistry (29.7%), physics (27%), biology/life sciences (21.6%), Earth science/geology (16.2%), and nature of science/scientific method (5.4%).

Classroom assessments were analyzed at the item level. The submitted assessments contained a total of 727 items. An item scoring rubric, based on Gotwals, Hokayem, and Song (2009), was used to code the cognitive demand of each item (see Figure 1). Following training on proper use of the scoring rubric, five researchers collaborated to code each item. All items were first coded by a rater working individually. Then, a team of two different raters worked together to assign each item a second, independent code based on the item scoring rubric. Interrater reliability between all groups of raters was established at 90%. Disagreements were resolved through discussion.

Cognitive demand of items was determined by analyzing both response format and cognitive processes needed to complete each item. This analysis of cognitive demand followed the pattern set by Mergendoller, Marchman, Mitman, and Packer (1988) in focusing on the structure of the problems presented to the students rather than the knowledge that students brought with them to solve those problems. This focus allowed for efficient analysis of a large number of items from many different teachers. Response formats were divided into two basic categories: verbal/visual restricted, which included various selected response type items such as multiple-choice, true and false, matching, labeling, and model interpretation, and verbal/visual extended, which consisted of constructed response type items including short answer, essay, graphing, drawing a picture, model creation, and performance items.

Response Format	**Process Used to Solve the Item**					
	Creating Explanations(CE) – content knowledge, other available info., and problem solving skills are used to create an explanation	*Interpreting Data (ID)* – analysis of tables, graphs, or other forms of data lead to the answer	*Application* – apply knowledge to a situation *Basic Application (BA)*	*Protocol Development (PD)*	*Reading Comprehension (RC)* – answers come from a reading passage	*Definition (D)* – define a term
Verbal/ Visual Restricted (T/F, Matching, MC, Labeling, Model Interpretation etc.)	10	2	2	2	1	1
Verbal/Visual Extended (Short Answer, Essay, Graphing, Draw a Picture etc.)	50	10	10	10	6	4

Figure 1. Item scoring rubric based on Gotwals et al. (2009).

Additional detail on cognitive demand came from examining specific processes required for successful completion of an item. Cognitive process categories included creating explanations, interpreting data, applying content knowledge, reading comprehension, and defining a term or concept. To create explanations students use content knowledge and other available information in conjunction with problem solving skills to generate a scientific explanation for a situation. In data interpretation items students analyze tables, graphs, or other sources of data, in the context of content knowledge. Application questions require students to use content knowledge in a novel situation. This category was further divided into basic application questions and protocol development questions due to the common request for students to apply content knowledge in the context of creating or analyzing experimental procedures. To solve reading comprehension questions students first read a relevant passage and then use information from the reading to answer the item. Definitional items are completed by selecting or supplying the meaning of a term or concept. Examples of each response format and process skill represented in the rubric are included in Appendix B.

While Gotwals et al. (2009) described that categories of problem solving processes did not represent an overall hierarchy of complexity, they noted that "being able to explain a scientific situation illustrates an ability to fuse content knowledge and complex reasoning" (p. 3). They concluded that the category of creating explanations was therefore the most demanding cognitive process skill represented. As is evident from the numerical values assigned to each square in the item scoring rubric, this study is working from that same framework; creating explanations provides teachers with the greatest opportunity to understand students' ability to perform scientific inquiry. Furthermore, it is assumed that there is additional hierarchical complexity in the response format and cognitive processes of each item type represented in the rubric. The

following paragraphs describe the rationale for assignment of numerical values to the item scoring rubric based on specific cognitive demands of each response format and cognitive process.

Researchers have regularly attempted to qualitatively describe the strengths and weaknesses of particular item types at providing information about specific target knowledge or skills (e.g. Stiggins, 2005). Numerical values for the item scoring rubric in this study were based on the assumption that some items are better suited to provide information about student understanding of, and skill in, using scientific inquiry. Further, it is assumed that having more items of a specific format will provide a better picture of student understanding than having fewer items of that same type. Items designed to measure scientific inquiry-reasoning skills are essential tools in meeting current demands of science education reform documents (Gotwals & Songer, 2006). This does not imply that items with a higher numerical value are always better than items assigned a low numerical value. It simply describes how some items, because of cognitive demands they place on students, are better suited to help teachers determine students' ability to use scientific inquiry.

It is a common misconception that multiple-choice and other selected response items can only be used to assess content knowledge, not reasoning or problem solving (Stiggins, Arter, Chappuis & Chappuis, 2007). Given this misconception, verbal/visual extended items generally make higher cognitive demands on students than verbal/visual restricted items because they require that the test taker produce answers from their own understanding rather than selecting the correct answer from a list of possible responses. Lack of prompts in items reduces the likelihood that guessing will lead to a correct answer (Stiggins et al., 2007). For this reason, all extended

items in the item scoring rubric are given a higher numerical value than corresponding restricted items.

The highest numerical value, for both extended and restricted categories (50 and 10 respectively), was assigned to items requiring students to create explanations. According to NRC (2000) the ability to create scientific explanations is an essential and foundational element of scientific inquiry and must be done in concert with the development of scientific content knowledge. To create explanations students must engage in significant higher-level thinking. Webb, Vesperman, and Ely (2005) use Depth of Knowledge (DOK) levels to describe the cognitive demands that can be inferred from item expectations. According to this classification scheme, items requiring creation of scientific explanation represent DOK level 3-strategic thinking or DOK level 4-extended thinking. DOK levels 3 and 4 require multi-step processes to solve them and often have more than one possible answer. Strategic thinking is abstract and complex. It could include such activities as supporting ideas with details and examples, creating research questions, formulating hypotheses, drawing conclusions, or creating models to explain scientific phenomena. Extended thinking requires students to make many connections either within or among content areas. It also requires students to select an appropriate approach to finding a solution from several possible alternatives. Examples of extended thinking activities include analyzing and synthesizing data available from multiple sources, using mathematical models to better understand a relationship or problem, and completing a guided or open inquiry investigation.

Additional cognitive complexity is introduced in items requiring the creation of explanation due to the significant transfer of knowledge needed to successfully complete such items. Bransford, Brown, and Cocking (1999) describe the transfer of learning from one context to a

novel but related context as an essential feature of quality education because it prepares students to be flexible when trying to solve a new problem. The ability of these items to assess knowledge transfer provides additional support for the high numeric value they are assigned on the item scoring rubric.

Data interpretation and application items receive the same numerical value on the scoring rubric (extended response = 10, restricted response = 2) because cognitive demands they place on students are nearly identical. When the response format of these items is restricted they typically require students to use a DOK level 2- skill/concept in order to complete the item. Skill/concept activities are generally simplistic multi-step problems requiring students to first recall some content knowledge and then use that knowledge to solve the problem. Such activities could include collecting data, identifying patterns, making observations, or organizing and interpreting data. Using the verbal/visual extended response format moves this type of item from a skill/concept activity to a DOK level 3- strategic thinking activity. Cognitive demands of strategic thinking items warrant a higher numerical value due to the increased item complexity resulting from the requirement that students explain their thinking (Webb et al., 2005). Additionally, data interpretation and application items require only minimal transfer and so can be viewed as less complex than items requiring the creation of scientific explanation.

Reading comprehension (RC) and definition (D) items were assigned the lowest numerical values in the item scoring rubric (extended response = 6 for RC and 4 for D, restricted response = 1) because of their relative cognitive simplicity. These items do not require knowledge transfer. Rather, they invite students to demonstrate rote memorization or completion of a set procedure. Such items are typically classified at a DOK level 1-recall and only require that students supply an answer rather than “solve” or “figure out” the problem. An answer is

automatically available to the student with the knowledge necessary to answer that item because no processing or application of information is required to solve recall items (Webb et al., 2005). An additional level of complexity is introduced when reading comprehension items use an extended response format. The higher numerical value assigned to extended response reading comprehension items is based on the fact that these items require students to engage in more than simple recall of information. This type of item is classified as a DOK level 2-skills/concepts item because it requires students to interpret information that they have read and make inferences based both on reading and their understanding of scientific content knowledge.

Analysis and Findings

This section describes results of methods used to establish teacher inquiry levels and overall assessment scores. It then describes the presence or absence of specific item types on each assessment. Finally, it presents information on two methods used to check correlation between teacher inquiry levels and assessment scores. Qualitative data from open-ended survey questions provide additional support and description of the quantitative findings.

Teacher Inquiry Levels, Assessment Scores, and Presence or Absence of Item Formats

A characterization of individual inquiry levels was determined using teacher responses to the Likert items that probed teachers' beliefs and pedagogical practices. Answer choices "never," "rarely (e.g. a few times a year)," "sometimes (e.g. once or twice a month)," "often (e.g. once or twice a week)," and "all or almost all science lessons" were assigned scores from -2 to 2. Items coded as detracting from an inquiry focus were assigned a negative multiplier so scores would reflect that more frequent participation in these activities was in opposition to methods supportive of scientific inquiry. Teachers were given an overall inquiry score based on the sum

of their scores on items describing pedagogical beliefs and practices supportive of an inquiry focus and items detracting from such efforts. Teacher inquiry scores ranged from -24 to 30 (mean = 0.01, SD = 10.9). Teachers were sorted into one of three groups using cut scores based on standard deviations: high (greater than 5.6), medium (between 5.5 and -5.5), and low (less than -5.6). Nearly equal numbers of teachers, high (32.5%), medium (37.4%), and low (30.1%), fell into each of the inquiry level groups.

Because only 30 of the original 83 survey respondents provided an example classroom assessment it was important to ensure that inquiry levels of teachers in this subgroup were not significantly different from inquiry levels of all survey respondents. A least squares means (LSM) test was used to compare inquiry levels of teachers that completed the survey and submitted an assessment (LSM = -0.933, p = 0.643) with those that only completed the survey (LSM = 0.547, p = 0.7179). These results indicate that there was not a significant difference in calculated inquiry levels between teachers completing only the survey and those completing the survey and submitting an assessment.

An overall assessment score for each teacher was calculated by summing the numeric values assigned to individual items using the item scoring rubric. Assessment scores for the three teachers submitting more than one assessment were determined by averaging scores from provided assessments. The wide range in assessment scores, low of 13 to high of 468 (mean = 101.1, SD = 89.1), was the combined result of extensive variation in the total number of items included on each assessment and differences in numerical values assigned to each item based on cognitive complexity. Natural log scores of overall assessment scores were used to reduce spread.

Teacher inquiry scores from responses to Likert items, inquiry levels based on cut scores, overall assessment scores, natural log assessment scores, and presence or absence of each item type from the item scoring rubric are summarized in Table 1.

Table 1

Teacher inquiry levels, assessment scores, and presence or absence of item types on test.

	Inquiry	Inquiry	Assessment	Log	Extended						Restricted					
Teacher	Score	Level	Score	Score	CE	ID	BA	PD	RC	D	CE	ID	BA	PD	RC	D
1	-20	L	99	4.595			X	X		X						X
2	-20	L	33	3.497								X	X			X
3	-19	L	27	3.296								X	X			X
4	-12	L	85	4.443						X	X	X	X		X	X
5	-11	L	97	4.575	X		X			X			X			X
6	-10	L	22	3.091									X	X	X	X
7	-10	L	50	3.912	X											
8	-9	L	73.5	4.297								X	X	X		X
9	-9	L	100	4.605	X					X						X
10	-7	L	13	2.565								X	X			X
11	-6	L	38.5	3.651								X	X		X	X
12	-3	M	150	5.011	X		X				X	X				
13	-2	M	102	4.625	X							X	X			X
14	-1	M	87[a]	4.466				X		X						
15	-1	M	40	3.689			X					X	X		X	X
16	1	M	106	4.663			X			X			X			X
17	1	M	33	3.497			X				X		X			X
18	2	M	82[a]	4.407	X		X	X		X						
19	3	M	111[a]	4.710			X			X						X
20	3	M	110[a]	4.700			X	X								
21	4	M	110[a]	4.700	X	X	X	X								
22	4	M	468	6.148	X	X	X			X			X		X	X
23	6	H	44	3.784									X			X
24	9	H	85[*]	4.443		X	X	X								
25	9	H	168	5.124			X			X						
26	11	H	287	5.659	X	X	X			X			X			X
27	12	H	80	4.382			X									
28	12	H	47[a]	3.850	X		X	X		X			X	X		
29	17	H	104	4.644	X		X			X						
30	18	H	180[a]	5.193	X		X									

Note. "X" indicates presence of that item type on a teacher's classroom assessment. [a] These assessments were laboratory activities or performance assessment rubrics.

Submitted classroom assessments frequently included some extended response and some restricted response items (43.3%). A smaller group included only extended items (33.3%) or only restricted items (23.3%). All but two assessments included items that required students to use at least two different types of cognitive processing skills to complete the assessment. Assessments that used only one item type and therefore required students to use only one cognitive process had a small total number of items (8 and 1).

Item types included most frequently in these assessments were definition items using a verbal restricted format (60%) and basic application items using an extended response format (60%). It was also common for these assessments to include basic application questions using a restricted response format (53.3%) or definitional items using an extended response format (43.3%). When assessments included items that required the creation of explanation, it was more common to employ an extended response format (40%) than a restricted response format (10%). The opposite was true of items requiring interpretation of data; a restricted response format (30%) was more common than an extended response format (13.3%). Protocol development items were included on nearly one third (30%) of the teacher submitted assessments. It was more common for these items to require an extended response (23.3%) than a restricted response (10%). Assessments employing reading comprehension items (16.7%) always used a restricted response format.

Correlation between Teacher Inquiry and Assessment Scores

Two similar methods were used to identify and examine a possible relationship between teacher inquiry and assessment scores. A linear regression and Pearson's correlation were calculated to look for connections between teacher inquiry scores and the natural log of overall

assessment scores (see Figure 2). A Tukey-Kramer pairwise analysis, in conjunction with an analysis of variance (ANOVA), was used to compare teacher inquiry levels (high, medium, and low-based on cut scores) to the natural log of assessment scores (see Table 2).

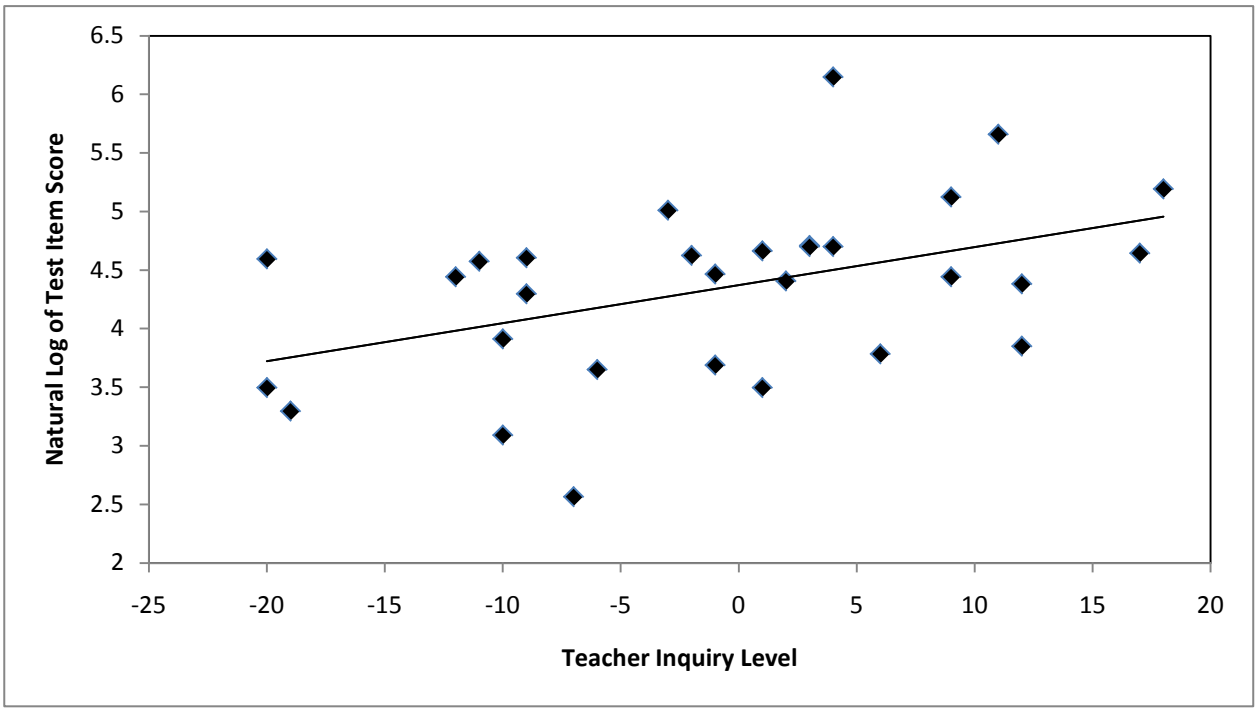


Figure 2. Graphical representation of the correlation between test item score and teacher inquiry level.

As indicated by the scatter plot of test item scores and teacher inquiry levels and the corresponding linear regression, results indicate a weak correlation ($r = 0.447$, $p = 0.0133$). Teachers that scored higher on use of inquiry in their classrooms also tended to score higher on their overall assessment scores. This demonstrates that teachers using inquiry methods in their teaching are more likely to also assess student understanding of scientific inquiry using cognitively complex items.

Table 2

Results of the Tukey-Kramer pairwise analysis.

Inquiry Level	LSM	SE	P(High	- Medium -	Low)
High	4.63	0.24		0.993	0.057
Medium	4.6	0.21	0.993		0.046
Low	3.87	0.21	0.057	0.046	

Note. The three groups were significantly different with a $p < 0.001$.

The results of the ANOVA and Tukey-Kramer analysis were similar to the results of the Pearson's correlation and linear regression. The ANOVA indicated a significant difference between the assessment scores for the three inquiry levels (DF = 2, p = 0.0262). Tukey-Kramer pairwise analysis suggested a significant difference between the least squares means of the natural log assessment scores in the following pairs of teacher inquiry levels: low inquiry level with medium inquiry level (p = 0.046) and low inquiry with high inquiry (p = 0.057). The least squares means of assessment scores for the medium inquiry level and the high inquiry level teachers were very similar and no significant difference was found (p = 0.993). These findings place emphasis on the fact that there is at least a moderate link between a teacher's use of inquiry methods in his or her pedagogical practices and the methods used to assess student understanding. Teachers in this study that demonstrated a greater inquiry focus in beliefs and described practices also used assessment items that allow them to more accurately assess student ability to participate in scientific inquiry.

Teacher Responses to Open-ended Survey Questions

Significant themes emerged during the coding of open ended questions. The following section describes each open-ended question including a list of all significant coding categories and frequency of category responses. Teacher respondents often discussed multiple themes in a

given response so that answers routinely fit simultaneously into several coding categories. Because of this, percentages of teachers including responses on all coding categories do not always sum to 100%.

Survey question 13 asked, “How would you define scientific inquiry?” Coding categories and frequency of responses are listed in Table 3. In response to this question, teachers generally expressed a fairly simplistic understanding of scientific inquiry. Respondents described scientific inquiry as, “asking a question and finding the answer” (Survey Respondent 2), “activities that pose a problem or question to the students and requires them to search the answer or come up with a solution” (Survey Respondent 30), and “guiding the students to find their own answers to scientific questions” (Survey Respondent 4). As these responses demonstrate, many teachers described how scientific inquiry requires students to ask and seek answers to questions (50.6%) or discover scientific principles using personal experiences (18.1%). Fewer teachers described a direct connection between scientific inquiry and the scientific method (30.1%) or experimentation (18.1%). The following example illustrates a more nuanced understanding of scientific inquiry which was less common in survey responses.

> Scientific inquiry is the asking of questions about a science topic, determining how best to answer those questions using the scientific method, developing a test, experiment or way to answer those questions, and then completing the test to see if one receives an answer or not. (Survey Respondent 76)

Table 3

Question 13 response categories. (n=83)

Coding Category	Frequency of Response	% of Teachers Including Response
Ask and Answer Questions	42	50.6
Scientific Method	25	30.1
Discovery	15	18.1
Experimentation	15	18.1
Hands-on	8	9.6
Student Centered	6	7.2
Uses Curiosity	6	7.2

In survey question 14 teachers were asked, "Do you think that it is important to use inquiry-based learning activities in your class? Why? or Why not?" Coding categories and frequency of responses for this question are listed in Table 4. Almost all respondents (96.4%) affirmed their belief that inquiry-based learning activities are important instructional tools. They explained that inquiry-based activities allow students to improve problem-solving skills (39.8%), increase student motivation (32.5%), lead to deeper, enduring student learning (27.7%), and strengthen students' understanding of the nature of science (NOS) (16.9%). Survey Respondent 34 described her reasons for including inquiry-based activities in the following way:

> I think it is important to use some inquiry-based learning activities in class because it gets the kids thinking about what is happening instead of the teacher just showing them. It also teaches them that a sloppy job of collecting data gives them bad data and sometimes they can't figure out the relationship. It gets the kids asking questions and even a desire to find out the answer. They are so used to the teacher answering all their questions.

Three respondents explained that they did not feel that inquiry-based activities were important for their students. One teacher explained that the benefits of inquiry activities do not outweigh the significant time costs needed to complete them. The others felt that such activities were not appropriate for their students because “students at the high school level don't know enough to be able to ask questions that are of value or that they are actually able to test” (Survey Respondent 66) or “because the maturity level of [students] is small and their . . . understanding [of] science related subjects very immature” (Survey Respondent 7).

Table 4

Question 14 response categories. (n=83)

Coding Category	Frequency of Response	% of Teachers Including Response
Yes	80	96.4
Problem Solving Skills	33	39.8
Student Interest and Motivation	27	32.5
Lasting Learning	23	27.7
NOS	14	16.9
No	3	3.6

Question 15 probed, “Do you think that it is important to assess your students’ ability to conduct scientific inquiry in your class? Why? or Why not?” Coding categories and frequency of responses are shown in Table 5. Nearly all teacher respondents (95%) expressed the belief that it is important to assess student scientific inquiry competency. Many respondents described general goals of assessment, including checking “to see if [students] understand what they are doing” (Survey Respondent 15) and the need to assess “everything that you think you are teaching students” (Survey Respondent 5), as reasons for assessing scientific inquiry (check understanding = 50%, assess what is taught = 6.3%). The connection between assessing

scientific inquiry and the larger goals of science education (8.8%), including the need to assess students understanding of NOS (25%), were also noted regularly in survey responses.

Notably, 12 of the respondents that expressed feeling that it is important to assess students' skills in scientific inquiry went on to describe reasons why they feel that actually doing this presents significant challenges. Survey Respondent 6 lamented, "In our test-driven education culture, it is difficult to find an assessment that properly tests inquiry." Other respondents noted that assessing these skills is "difficult to do in a large classroom situation" (Survey Respondent 12) and that it is "more difficult and time consuming compared to worksheets or [multiple-choice] questions" (Survey Respondent 77).

Table 5

Question 15 response categories. (n=80)

Coding Category	Frequency of Response	% of Teachers Including Response
Yes	76	95.0
Check Understanding	40	50.0
NOS	20	25.0
Goal of Science Education	7	8.8
Assess What is Taught	5	6.3
No	4	5.0

Question 17 queried, "Do you use inquiry-based learning activities in your class? If yes, give an example of a typical inquiry activity that you use in your class. If not, why not?" Table 6 shows coding categories and frequency of responses for this question. A large majority of respondents (yes = 78.5%, sometimes = 15.2%) described using inquiry-based learning activities with their students. When sufficient details were included in the response, "typical inquiry

activities" were coded using a priori codes for inquiry levels of classroom activities described by Chinn and Malhotra (2001) and Windschitl (2003).

Confirmation activities were described more often than any other type of inquiry investigation (29.1%). In confirmation activities students follow step-by-step instructions to verify a stated scientific principle. Survey Respondent 80 described the following confirmation activity:

> To teach density I have several cubes of the same size, but made of different material[s]. The students measure the mass and volume and then order the cubes from least dense to most dense. Then they match the cube with a layer of the earth.

Structured inquiry activities, which provide students with a question and the procedures needed to answer the question, were also fairly common (24.1%). A typical structured inquiry activity was described in the following way:

> The students are given a bunch of reactants and asked to write a complete equation and predict which compounds when mixed will precipitate and what the precipitate will be. Then they test their answers in lab and discuss the outcomes. (Survey Respondent 50)

Guided inquiry activities (17.7%) and open inquiry activities (3.8%) were described less frequently by survey respondents. Guided inquiry activities are distinguished by student autonomy to design the experimental procedures, select variables, and explain the results in order

to answer questions put forward by the teacher. These characteristics are emphasized in the following example of a guided inquiry activity:

> I do a physical separation lab in class where each student must come up with methods to separate salt, sand, sawdust, and iron filings. They create their own material lists, procedure steps, data table, flow chart, and show proof of separation by placing all the items in separate plastic bags. This is very challenging for an 8th grader. (Survey Respondent 3)

The following example of an open inquiry activity highlights how this type of activity gives students additional freedom to formulate research questions, design appropriate experiments, analyze and interpret data, and share findings with a larger research community.

> Students monitor their river and the bosque (forest) surrounding the river on a monthly basis as an assessment of the ecosystem's overall health. Students' questions guide our study and their analysis of the data which then gets passed on to university and state game and fish departments. (Survey Respondent 76)

Sufficient information was not provided to be able to determine the inquiry level of some reported classroom activities (17.7%). Additionally, 6.3% of respondents claimed to not use any inquiry-based learning activities with their students.

Table 6

Question 17 response categories. (n=79)

Coding Category	Frequency of Response	% of Teachers Including Response
Yes	62	78.5
Sometimes	12	15.2
Confirmation	23	29.1
Structured	19	24.1
Guided	14	17.7
Open	3	3.8
Not Enough Detail to Tell	14	17.7
No	5	6.3

Survey question 18 asked, “Do you assess your students’ ability to conduct scientific inquiry? If yes, give an example of how you do this in your class. If not, why not?” Coding categories and frequency of responses are shown in Table 7. Nearly 80% of respondents reported using some method to assess student competency in scientific inquiry. Only a small fraction of respondents (5.1%) described using a formal test as a means of assessing this element of science. Nearly half (43%) agreed with the idea that, “lab reports allow students to communicate their ability to conduct scientific inquiry” (Survey Respondent 26). About one quarter (24.1%) of respondents described how careful observation and probing questioning during inquiry activities provides the most valuable feedback on students’ inquiry skills. Survey Respondent 26 offered the following insight on how extended observation and questioning can be used to provide feedback, to teacher and student, on student understanding of scientific inquiry.

> I assess my students' ability to conduct scientific inquiry mostly with whiteboarding exercises. These can be formative assessments as well as summative. Students are

presented with a problem and then are given time to collect data about the problem. Then, as a class, each group's data is presented on a whiteboard to the class. The class and I ask questions about their procedure, data, evaluation and conclusion to which the group presenting must respond. This method is useful for finding out how students think about scientific processes and how they carry them out.

Twenty percent of survey respondents explained that they do not assess students' scientific inquiry capabilities. Reasons for not assessing student competency in scientific inquiry included "students' lack [of] science lab skills" (Survey Respondent 34), the "limited time [students] have in the classroom" (Survey Respondent 4), a shortage of "effective way[s] to assess students' ability to conduct scientific inquiry" (Survey Respondent 6), and the feeling that this type of assessment is "more difficult to do" (Survey Respondent 63) than assessments of content knowledge alone.

Table 7

Question 18 response categories. (n=79)

Coding Category	Frequency of Response	% of Teachers Including Response
Yes	63	79.7
Lab	34	43.0
Observation/Questioning	19	24.1
Performance Assessment	8	10.1
Rubric	5	6.3
Test	4	5.1
No	16	20.3

The final open-ended item, survey question 20, asked teachers, “What, if any, influence do state or district mandated science accountability assessments have on the teaching and assessment methods that you use in your classroom?” Table 8 includes coding categories and response frequencies for this item. Many survey respondents expressed the feeling that large-scale accountability assessments in science are impacting their classroom teaching and assessment practices (major impact = 31%, moderate impact = 46.4%). One teacher described the impact of large-scale accountability assessments in this way:

> We follow the state core. We have altered what we teach to be in line with the state standards and the concepts that are on the end of level test. I try to assess the standards. I use a number of methods: common assessments based on the core standards, written formal lab reports and unit tests and quizzes. (Survey Respondent 14)

This quote explains how accountability assessments are impacting pedagogy and assessment practices. Many teachers responding to this survey also described a connection between accountability assessments and the state science curriculum on which they are based. Nearly half of respondents (49.3%) mentioned the goal of focusing instruction on science core curriculum with which state and district assessments are aligned.

Over one fourth of respondents (26.8%) reported that they used assessment practices employed on large-scale assessments, including item formats, wording, and electronic test administration. Survey Respondent 44 stated, “I spend a lot of time making the students comfortable with the format of the state test and how to read and understand what they are asking.” Respondents (18.3%) explained that they regularly prepare students for large-scale

accountability assessments in science by reviewing material that they feel is likely to show up on these tests. A sizeable group of survey respondents (22.5%) expressed the feeling that state and district science accountability assessments were not impacting classroom assessment or pedagogy.

Table 8.

Question 20 response categories. (n=71)

Coding Category	Frequency of Response	% of Teachers Including Response
Major	22	31.0
Moderate	33	46.4
Teach Core	35	49.3
Assess Like State	19	26.8
Review/Test Preparation	13	18.3
None	16	22.5

Discussion

The initial questions guiding this research were:

1. Can a teacher's beliefs and practices regarding inquiry teaching methods be correlated with his/her assessment practices?
2. What item types are most commonly employed by teachers that use an inquiry pedagogy?
3. What assessment strategies do teachers describe to assess scientific inquiry?

The paragraphs that follow summarize what has been learned about each of these questions as a result of this study. Limitations of this study and possible areas of future research are also outlined.

Correlation between Teacher Inquiry Levels and Classroom Assessment Practices

Science teachers participating in this study employed various levels of inquiry teaching methods and learning activities within their classrooms. Using two methods, this study found a statistically significant correlation between teacher inquiry implementation and classroom assessment practices. Teachers that use inquiry teaching methods more frequently also employed more cognitively complex items on classroom assessments. This correlation, though modest, indicates that teachers are able to simultaneously focus on inquiry in pedagogical and assessment practices.

Bol and Strage (1996) explained science educators routinely adopt instructional goals focusing on critical thinking and scientific inquiry, but they rarely assess these capabilities in students. This study indicates that teachers are working to improve alignment between instructional goals and assessment practice. However, a gap still exists between teachers' use of inquiry focused pedagogy and assessment of student understanding of scientific inquiry.

Almost all survey respondents reported that it was important to include inquiry-based activities during classroom instruction (96.4%) and to assess student competency in scientific inquiry (95%). An equally large proportion of survey respondents reported using inquiry-based activities (95%). However, only 80% of respondents claimed to assess student competence in scientific inquiry.

In a case study examining assessment beliefs and practices of three high school science teachers, Aydeniz (2007) describes that a key problem in implementing assessment reform has been individual teachers' naïve pedagogical content knowledge. Similarly, this study both supports and expands upon this idea. Even among teachers reporting that it is important to assess

students' scientific inquiry competency, many survey respondents spontaneously described barriers to such assessments. Factors limiting their ability to assess student inquiry competencies included restricted teacher knowledge of how to create such assessments, and a dearth of existing assessment materials for use either as models or for direct implementation. Additionally, teachers describe limited time and lack of classroom space and resources make it challenging for them to implement the types of assessments that would give the best feedback on student competency in scientific inquiry.

Item Types used by Teachers with an Inquiry Focus

Lane (2004) argued that item types used on science assessments should focus students and teachers towards practicing problem-solving and reasoning skills. We found that teachers with an inquiry focus are moving towards Lane's ideal in teacher-created classroom assessments. The correlation found in our study indicates that teachers reporting an inquiry focus in pedagogy are also inclined to use cognitively complex items when assessing students. Cognitively complex items provide clearer feedback about scientific inquiry competency because they require the same sets of cognitive skills as scientific inquiry.

Historical data indicate that different item formats do not necessarily measure different constructs (Hollingworth et al., 2007), but work on item format and cognitive complexity reported that inclusion of multiple item formats may tap different levels of student cognition (Martinez, 1999). Researchers recently created a standards-based assessment which contained both multiple-choice and open-ended questions aligned to a state core curriculum. The authors reported mixed results. Sometimes it appeared that multiple-choice and open-ended items were measuring different constructs and other times they seemed to be measuring the same construct

(Hollingworth et al., 2007). Despite confusion in the literature, teachers in this study regularly used multiple item types and response formats on classroom assessments. All but two of the teacher-created assessments included multiple item types and 43.3% included both extended response and restricted response formats. Teachers from this study were inclined to use a variety of item types to assess students.

Respondents reported many challenges in creating assessments that move beyond assessing content knowledge. Teachers in this study described the need for additional methods to effectively assess inquiry-based teaching and the resulting changes in student understanding of scientific inquiry. Researchers have described similar problems for large-scale assessments (Lane, 2004). Orpwood (2001) stressed that new assessment approaches need to be designed and implemented to keep up with the changing curricular emphases. Recognizing this problem in science, researchers have created item types and assessment systems designed to help differentiate between students' content knowledge and their inquiry-reasoning abilities (Liu et al., 2008; Gotwals & Songer, 2006; Toth, Suthers, & Lesgold, 2002; Wilson & Sloane, 2000; Zachos et al., 2000). However, teachers need access to these assessment tools and item types both for direct implementation in classrooms and as models for future classroom-level item development.

Assessment Methods Teachers Use to Assess Student Competency in Scientific Inquiry

This study invited teachers to submit a typical classroom assessment that measured student understanding of scientific inquiry. However, it is clear from the collected assessments that traditional assessments are not the only method used to collect this type of information. Over one fourth of survey respondents (26.7%) sent an open laboratory investigation or performance

assessment scoring rubric rather than a traditional restricted assessment. Furthermore, in response to open-ended question 17, only 5.1% of respondents indicated that they use traditional tests to assess inquiry competency. Alternative assessment practices including labs (43%), questioning and observation (24.1%), performance assessments (10.1%), and rubrics (6.3%) were all more commonly mentioned than traditional testing methods.

An investigation by Lawrenz, Huffman and Welch (2001) attempted to compare the fairness of several assessment types including: a multiple-choice test, an open ended written test, a laboratory test, and a full laboratory investigation. Varied performance on different test formats indicated that different item types are measuring different constructs. In particular, lab tests and full lab investigations seemed to be measuring something different than multiple-choice or open-ended questions. Respondents in our study seemed to intuitively agree with the idea that laboratory settings measure a different construct than more traditional item types. Forty-three percent of survey respondents indicated that laboratory investigations represent a preferred method of assessing scientific inquiry competency in students.

Lorsbach et al. (1992) found that oral examinations were effective tools for assessing scientific inquiry. When teachers allowed students to express understanding orally they were able to use questioning techniques to probe student competency in scientific inquiry. The results of our study also indicate that teachers regularly use observation and questioning techniques to measure inquiry skills. Nearly one quarter of survey respondents reported assessing student competence in scientific inquiry through extended observation and questioning.

Limitations

The small overall sample size (N = 83) and even smaller number of teachers submitting a classroom assessment document (n = 30) represent potential limitations to this study. It is possible that the fairly low overall response rate, to both the initial survey and the follow-up invitation to submit a teacher-made assessment, introduced sampling bias. Respondents may have held stronger opinions and beliefs about classroom assessment than a random sample of science teachers. Further, respondents were almost all from the same state (96.4%), had been teaching for a long time (mean = 14.9, SD = 7.62), and the majority (65%) reported having an advanced degree in science or science education which may indicate that they have completed more training in scientific inquiry and assessment than the average science teacher. These factors may limit the transferability of the findings of this study to a broader context.

Additional limitation in this study may be the result of high variation in overall assessment length as measured by the total number of items. This variation made it challenging to compare assessments from different teachers. It was difficult to give an average item score that accurately reflected the cognitive complexity of all assessment items. Instead assessments were given an overall score, but this score does not take into account the time that students were given to work on a particular assessment.

Future Research

The modest correlation between teacher inquiry level and use of cognitively complex assessment items reported in this study clarifies how science teachers are simultaneously meeting the goals of inquiry and assessment reform movements. However, additional research is needed. Future research should focus on understanding what forces are most influential in shaping

assessment practices of science teachers, including development of item and assessment formats used at the classroom level. Additionally, researchers should create easy to use scientific inquiry assessment tools that can be employed regularly by teachers, both through direct implementation and as models for teacher-based assessment development, to assess student competency in scientific inquiry. In the future researchers may be more successful in collecting data on classroom assessment practices from teachers if they link data collection efforts to professional development opportunities for teachers.

References

American Association for the Advancement of Science (2006). *Project 2061: benchmarks for science literacy-a tool for curriculum reform.* Retrieved February 13, 2009, from http://www.project2061.org/publications/bsl/default.htm

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

Anderson, R. D. (2007). Inquiry as an organizing theme for science curricula. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 807-830). Mahwah, N.J. : Lawrence Erlbaum Associates.

Aydeniz, M., (2007). Understanding the challenges to the implementation of assessment reform in science classrooms: a case study of science teachers' conceptions and practices of assessment. *Retrieved from ProQuest Digital Dissertations.* (AAT 3263843)

Bell, B. (2007). Classroom assessment of science learning. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 965-1006). Mahwah, N.J. : Lawrence Erlbaum Associates.

Bell, L. R., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction: assessing the inquiry level of classroom activities. *The Science Teacher. 72*(7), 30-33.

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education, 80*(2), 145-163.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: brain, mind, experience, and school,* (Ed.). Washington, D.C.: National Academy Press.

Britton, E. D., & Schneider, S. A. (2007). Large-scale assessments in science education. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 1007-1042). Mahwah, N.J. : Lawrence Erlbaum Associates.

Bybee, R. (2000). Teaching science as inquiry. In J. Minstrell, & E. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 20–46). Washington, DC: American Association for the Advancement of Science.

Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175- 218.

Champagne, A. B., Kouba, V. L., & Hurley, M. (2000). Assessing inquiry. In J. Minstrell, & E. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 447-470). Washington, DC: American Association for the Advancement of Science.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047-1085.

Duschl, R., & Gitomer, D. H. (1997). Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms. *Educational Assessment*, *4*, 37-73.

Enger, S. K. (1997). Linking instruction and assessment in science: Science learning opportunities and student performance on a set of open-ended science questions. Presented at the Annual Meeting of the Mid-South Educational Research Association (November 12-14, 1997).

Germann, P., Haskins, S.,& Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science Teaching*, *33*(5), 475–499.

Gotwals, A. W., Hokayem, H., & Song, T. (2009). *What types of cognitive demands do standardized testing items place on students?* Paper presented at the Annual Meeting of the National Association of Research in Science Teaching (April 17-21, 2009).

Gotwals, A. W., & Songer, N. B. (2006). Measuring students' scientific content and inquiry *reasoning*. Paper presented at the international conference on learning sciences, Bloomington, Indiana.

Hanegan, N. & Friden, K. (2009). Authentic and simulated inquiry: Teachers reflect what is modeled. *School Science and Mathematics Journal.* 109 (2) 110-134.

Herron, M. D. (1971). The nature of scientific inquiry. *School Review, 79*, 171–212.

Hollingworth, L., Beard, J. J., & Proctor, T. P. (2007). An investigation of item type in a standards-based assessment. *Practical Assessment, Research & Evaluation*, 12(18).

Horizon Research, Inc. (2000). *2000 National Survey of Science and Mathematics Education Science Questionnaire.* Retrieved March 4, 2009, from http://www.horizon-research.com/instruments/2000survey/tq_sci.php

Johnson, C., & Hanegan, N. (2006). No Child Left Behind-What does this mean for middle school teachers? *Science Scope*. 30 (3) 12-16.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher.* 33 (7), 14-26.

Lane, S. (2004). Validity of high-stakes assessment: are students engaged in complex thinking? *Educational Measurement: Issues and Practice*, *23*(3), 6-14.

Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education, 85*(3), 279-290.

Le Floch, K. C., Martinez, F., O'Day, J., Stecher, B., Taylor, J., & Cook, A. (2007). *State and local implementation of the no child left behind act volume III-accountability under NCLB: interim report*. Retrieved July 23, 2008, from http://www.ed.gov/rschstat/eval/disadv/nclb-accountability/index.html

Liu, O. L., Lee, H., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, 13(1), 33-55.

Lorsbach, A. W., Tobin, K., Briscoe, C., & LaMaster, S. U. (1992). An interpretation of assessment methods in middle school science. *International Journal of Science Education*, *14*(3), 305-317.

Martinez, M. (1999). A comparison of multiple-choice and constructed figural response items. *Journal of Educational Measurement*. 28(2), 131-145.

Mergendoller, J. R., Marchman, V. A., Mitman, A. L., & Packer, M. J. (1988). Task demands and accountability in middle-grade science classes. *Elementary Science Journal*, 88(3), 251-265.

Miller, M. D., Lynn, R. L. & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching.* (10th ed.). Upper Saddle River, New Jersey and Columbus, Ohio: Merrill.

Morrison, S. E. (2008). *Partial inquiry implementation affects questioning behaviors during cognitive activities.* Unpublished master's thesis, Brigham Young University, Utah.

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

National Research Council. (2000). *Inquiry and the national science education standards.* Washington, DC: National Academy Press.

No Child Left Behind Act of 2001, Public Law 107-110, 107th Cong., 115 Stat. 1425. (2002).

Orpwood, G. (2001). The Role of Assessment in Science Curriculum Reform. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 135-151.

Parx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in middle grades: assessment of learning and urban systemic reform. *Journal of Research and Science Teaching*, 41 (10), 1063-1080.

Schwab, J. (1962). The teaching of science as enquiry. In J. J. Schwab & P. F. Brandweine (Eds.), *The teaching of science*. (pp. 1-103) Cambridge, MA: Harvard University Press.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment and science. *Applied Measurement and Education,4*(4), 347-362.

Songer, N. B., & Gotwals, A. W. (2004). *What constitutes evidence of complex reasoning in science?* Paper presented at the proceedings of the 6th international conference on learning sciences.

Stemler, Steve (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Retrieved April 5, 2010 from http://PAREonline.net/getvn.asp?v=7&n=17

Stiggins, R. J. (2005). *Student-involved assessments for learning 4th edition*. Upper Saddle River, NJ: Merril/Prentice Hall.

Stiggins, R. J., Arter, J., Chappuis, J., & Chappuis, S. (2007). *Classroom assessment for student learning: doing it right-using it well*. Upper Saddle River, NJ: Merril/Prentice Hall.

Stoddart, T., Abrams, R., Gasper, E., & Canaday, D. (2000, December). Concept maps as assessment in science inquiry learning - a report of methodology. *International Journal of Science Education*, *22*(12), 1221-1246.

Tafoya, E., Sunal, D.,& Knecht, P. (1980). Assessing inquiry potential: A tool for curriculum decision makers. *School Science and Mathematics, 80,* 43–48.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). “Mapping to know”: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education, 86*(2), 264-286.

Trester, T.R. & Jones, M. G. (2003). Relationships between inquiry-based teaching and physical science standardized test scores. *School Science and Mathematics*, 103(7), 345-350.

Vogler, K. E., (2002). The impact of high stakes, state-mandated student performance assessment on teachers’ instructional practices. *Education, 123*(1), 39- 55.

Webb, N. L., Vesperman, B., & Ely, R. (24 July 2005). *Web alignment tool*. Wisconsin center for education and research. University of Wisconsin- Madison. Retrieved from http://wat.wceruw.org/Tutorial/SciDOKDef.aspx

Wideen, M. F., O’Shea, T. Pye, I. & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education, 22*(4), 428- 444.

Wilson, M., & Sloane, K. (2000). From principles to practice: an embedded assessment system. *Applied Measurement in Education*, 13(2), 181–28.

Windschitl, M. (2003). Inquiry projects in science teacher education: what can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education, 87(*1), 112-143.

Yeh, S. S. (2001). Tests worth teaching to: constructing state-mandated test that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.

Zachos, P., Hick, T. L., Doane, W. E. J., & Sargent, C. (2000). Setting theoretical and empirical foundations for assessing scientific inquiry and discovery in educational programs. *Journal of Research in Science Teaching, 37*(9), 938-962.

Appendix A – Teacher Survey

Modified from Horizon Research, Inc. (2000) National Survey of Science and Mathematics Education Science Questionnaire

Assessing Scientific Inquiry

Consent to Participate

Consent to be a Research Subject - Survey

Introduction
This research is being conducted by Dr. Nikki Hanegan and Mr. Adam Mitchell at Brigham Young University to determine how teachers' attitudes, beliefs, and application of inquiry science teaching methods influence their assessment practices. You have been selected to participate in this study because you currently teach one or more science class to students in grades 7 to 12. If you are no longer teaching, please send an email to mr.adammitchell@gmail.com and we will remove you from this study.

Procedures
As a research participant you will be asked to complete a survey describing your beliefs and goals pertaining to both inquiry teaching and assessment. Additionally, you will be invited to submit an example of an assessment document (e.g. a unit test) that you used this year.

Risks/Discomforts
This research poses minimal risks to participants. It is possible that you will experience slight emotional discomfort when answering survey questions or submitting an assessment used in your classroom. However, no questions in the survey are intended to be personally intrusive or deal with sensitive material.

Benefits
Individuals participating in this research will likely benefit from the opportunity that the survey questions provide to reflect on their teaching and assessment practices. It is likely that this reflection will increase your awareness of personal beliefs and practices in regards to assessment and inquiry teaching.

Confidentiality
Survey results and example assessment documents will be stored on BYU campus in a secure research laboratory. All data will be stored electronically on password-protected computers. Only the principal investigator and co-investigator on this project will have access to this data. Personal identifiers will not be attached to the survey results or the submitted assessment documents. Pseudonyms will be used to code and label electronic files and also in all work published as a result of this research.

Compensation
Research participants will receive no compensation for involvement in this study

Participation
Participation in this research study is strictly voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your standing with the University or researchers.

Questions about the Research
If you have any questions regarding this study, you may contact Adam Mitchell at 801-422-5468 or through e-mail at mr.adammitchell@gmail.com.

Questions about your Rights as Research Participants
If you have questions about this research or your rights, please contact IRB Chair, Dr. Christopher Dromey at 801-422-6461 or by e-mail at dromey@byu.edu.

Signature
I have read, understood, and received a copy of the above consent form and desire of my own free will to participate in this study.

Assessing Scientific Inquiry

*** 1. Please acknowledge that you have read and understand the above information by typing your initials and the date into the boxes below.**

Initials:

Date:

Assessing Scientific Inquiry

Demograhpic Information

2. Indicate your sex:

- ◯ Female
- ◯ Male

3. Are you: (Mark all that apply.)

- ☐ American Indian or Alaskan Native
- ☐ Asian
- ☐ Black or African-American
- ☐ Hispanic or Latino
- ☐ Native Hawaiian or other Pacific Islander
- ☐ White

4. How many years have you taught at the K-12 level including this school year?

Total Years = _____

5. What science class(es) do you currently teach? (List all.)

_____

6. What is your average class size this year?

Average Class Size = _____

7. Do you have each of the following degrees? (Mark all that apply.)

	Yes
Bachelors	☐
Masters	☐
Doctorate	☐

8. Please indicate the subject(s) for each of your degrees. (Mark all that apply.)

	Bachelors	Masters	Doctorate
Biology/Life Science	☐	☐	☐
Chemistry	☐	☐	☐
Earth/Space Science	☐	☐	☐
Physics	☐	☐	☐
Science Education (any science discipline)	☐	☐	☐
Mathematics/Mathematics Education	☐	☐	☐
Elementary Education	☐	☐	☐

Other (please specify)

9. How many total students attend the school at which you teach?

Total Students =

10. How would you classify the community in which your school is located?

- ◯ Urban
- ◯ Suburban
- ◯ Rural

11. What is the racial composition of your school? (Use numbers without a percent sign. The total of all groups together should equal 100.)

Percent American Indian or Alaskan Native

Percent Asian

Percent Black or African-American

Percent Hispanic or Latino

Percent Native Hawaiian or other Pacific Islander

Percent White

12. What percentage of the student population at your school qualifies for free or reduced lunch?

Percent Free/Reduced Lunch

Assessing Scientific Inquiry

Teacher Beliefs

13. How would you define scientific inquiry?

14. Do you think that it is important to use inquiry-based learning activities in your class? Why? or Why not?

15. Do you think that it is important to assess your students' ability to conduct scientific inquiry in your class? Why? or Why not?

16. Think about your teaching in one science class during the entire course. How much emphasis does each of the following student objectives receive? (Mark one answer for each line.)

	None	Minimal Emphasis	Moderate Emphasis	Heavy Emphasis
1. Increase students' interest in science	○	○	○	○
2. Learn basic science concepts	○	○	○	○
3. Learn important terms and facts of science	○	○	○	○
4. Learn science process/inquiry skills	○	○	○	○
5. Prepare for further study in science	○	○	○	○
6. Learn to evaluate arguments based on scientific evidence	○	○	○	○
7. Learn how to communicate ideas in science effectively	○	○	○	○
8. Learn about the applications of science in business and industry	○	○	○	○
9. Learn about the relationship between science, technology, and society	○	○	○	○
Learn about the history and nature of science	○	○	○	○
10. Prepare for standardized tests	○	○	○	○

Assessing Scientific Inquiry

Teacher Practices

17. Do you use inquiry-based learning activities in your class? If yes, give an example of a typical inquiry activity that you use in your class. If not, why not?

18. Do you assess your students' ability to conduct scientific inquiry? If yes, give an example of how you do this in your class. If not, why not?

19. Do the students you teach participate in accountability assessments for science (i.e. state or district mandated end-of-level tests)?

- ◯ Yes
- ◯ No

20. What, if any, influence do state or district mandated science accountability assessments have on the teaching and assessment methods that you use in your classroom?

21. About how often do you do each of the following in your science instruction? (Mark one answer for each line.)

	Never	Rarely (e.g. a few times a year)	Sometimes (e.g. once or twice a month)	Often (e.g. once or twice a week)	All or almost all science lessons
11. Introduce content through formal presentations	◯	◯	◯	◯	◯
12. Pose open-ended questions	◯	◯	◯	◯	◯
13. Engage the whole class in discussions	◯	◯	◯	◯	◯
14. Require students to supply evidence to support their claims	◯	◯	◯	◯	◯
15. Ask students to explain concepts to one another	◯	◯	◯	◯	◯
16. Ask students to consider alternative explanations	◯	◯	◯	◯	◯
17. Allow students to work at their own pace	◯	◯	◯	◯	◯
18. Help students see connections between science and other disciplines	◯	◯	◯	◯	◯
Assign science homework	◯	◯	◯	◯	◯
19. Read and comment on the reflections students have written, e.g., in their journals	◯	◯	◯	◯	◯

22. About how often do students in your science class take part in the following types of activities? (Mark one answer for each line.)

	Never	Rarely (e.g. a few times a year)	Sometimes (e.g. once or twice a month)	Often (e.g. once or twice a week)	All or almost all science lessons
20. Listen and take notes during presentation by teacher	◯	◯	◯	◯	◯
21. Watch a science demonstration	◯	◯	◯	◯	◯
Work in groups	◯	◯	◯	◯	◯
22. Read from a science textbook in class	◯	◯	◯	◯	◯
23. Read other (non-textbook) science-related materials in class	◯	◯	◯	◯	◯
24. Do hands-on/laboratory science activities or investigations	◯	◯	◯	◯	◯
25. Follow specific instructions in an activity or investigation	◯	◯	◯	◯	◯
26. Design or implement their own investigation	◯	◯	◯	◯	◯
27. Participate in field work	◯	◯	◯	◯	◯
28. Prepare written science reports	◯	◯	◯	◯	◯

23. About how often do students in your science class take part in the following types of activities? (Mark one answer for each line.)

	Never	Rarely (e.g. a few times a year)	Sometimes (e.g. once or twice a month)	Often (e.g. once or twice a week)	All or almost all science lessons
29. Record, represent, and/or analyze data	◯	◯	◯	◯	◯
30. Write reflections (e.g., in a journal)	◯	◯	◯	◯	◯
Prepare written science reports	◯	◯	◯	◯	◯
31. Make formal presentations to the rest of the class	◯	◯	◯	◯	◯
32. Work on extended science investigations or projects (a week or more in duration)	◯	◯	◯	◯	◯
33. Use computers as a tool (e.g., spreadsheets, data analysis)	◯	◯	◯	◯	◯
34. Use mathematics as a tool in problem-solving	◯	◯	◯	◯	◯
Take field trips	◯	◯	◯	◯	◯
35. Watch audiovisual presentations (e.g., videotapes, CD-ROMs, videodiscs, television programs, films, or filmstrips)	◯	◯	◯	◯	◯

Assessing Scientific Inquiry

24. How often do you assess student progress in science in each of the following ways? (Mark one answer for each line.)

	Never	Rarely (e.g. a few times a year)	Sometimes (e.g. once or twice a month)	Often (e.g. once or twice a week)	All or almost all science lessons
36. Conduct a pre-assessment to determine what students already know	◯	◯	◯	◯	◯
37. Observe students and ask questions as they work individually	◯	◯	◯	◯	◯
Observe students and ask questions as they work in small groups	◯	◯	◯	◯	◯
Ask students questions during large group discussions	◯	◯	◯	◯	◯
38. Use assessments embedded in class activities to see if students are "getting it	◯	◯	◯	◯	◯
39. Review student homework	◯	◯	◯	◯	◯
40. Review student notebooks/journals	◯	◯	◯	◯	◯

Assessing Scientific Inquiry

25. How often do you assess student progress in science in each of the following ways? (Mark one answer for each line.)

	Never	Rarely (e.g. a few times a year)	Sometimes (e.g. once or twice a month)	Often (e.g. once or twice a week)	All or almost all science lessons
41. Review student portfolios	◯	◯	◯	◯	◯
42. Have students do long-term science projects	◯	◯	◯	◯	◯
43. Give predominantly short-answer tests (e.g., multiple choice, true/false, fill in the blank)	◯	◯	◯	◯	◯
44. Give tests requiring open-ended responses (e.g., descriptions, explanations)	◯	◯	◯	◯	◯
Grade student work on open-ended and/or laboratory tasks using defined criteria (e.g., a scoring rubric)	◯	◯	◯	◯	◯
45. Have students assess each other (peer evaluation)	◯	◯	◯	◯	◯

Assessing Scientific Inquiry

Thank You

26. Thank you for the time you took to complete this survey. Would you be willing to participate in a follow-up interview for this study?

- ◯ Yes
- ◯ No

Appendix B – Sample Items from Assessments

Verbal/Visual Restricted

Creating Explanations

Models are used to represent atoms. The model we use today is not the same as the model used 200 years ago. What is a possible explanation for the change in the atomic model?

a. Scientists in the past couldn't see atoms, but today scientists have seen the atom and have developed a 100% correct, perfect, unchangeable model of the atom. There is nothing left to learn about atoms.
b. Scientists conducted further experiments and re-did the old experiments and were able to make a more modern, completely correct, perfect model.
c. Scientists in the past didn't have modern equipment so they were completely wrong. Today's scientists had to start the new atomic theory from scratch.
d. Scientists built on old knowledge and conducted further investigations on atoms to come up with the model we use today. The conclusions and the model may change as new discoveries are made.

Interpreting Data

Use the information in the table to answer the next question.

Properties of Four Substances

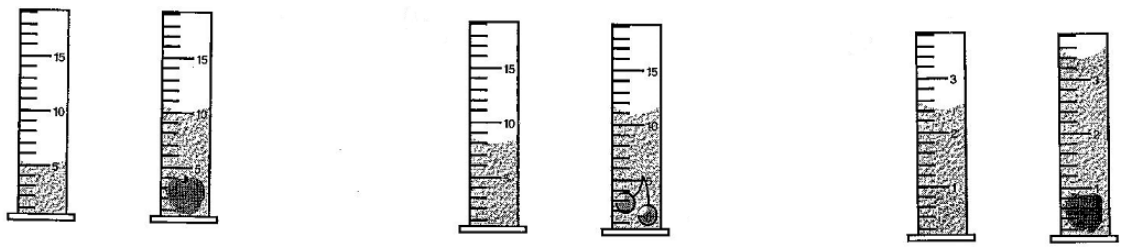
Substance	**Density**	**Phase at Room Temperature**	**Reaction with Water**	**Reaction to Flame**
Hydrogen	0.00009 g/ml	Gas	None	Burns explosively
Sodium	0.97 g/ml	Solid	Violent bubbling reaction	Burns explosively
Carbon	2.2 g/ml	Solid	None	Burns slowly
Argon	0.002 g/ml	Gas	None	none

Which substance showed no chemical change?

a. hydrogen
b. sodium
c. carbon
d. argon

Basic Application

Given the following graduated cylinders, what is the volume of the object?



_____ _____ _____

Protocol Development

A student wanted to find out which brand of popcorn had the least amount of unpopped kernels so he did an experiment where he tested three different brands. He popped one bag of each brand in the same microwave for the same amount of time and then counted the number of unpopped kernels in each bag.

For the experiment described above, what is most likely the student's hypothesis?

a. If I popped different brands of popcorn, the best brand will have the fewest unpopped kernels.
b. If I popped popcorn for different amounts of time, the longest time will have the fewest unpopped kernels.
c. All brands of popcorn are the same.
d. Which brand has the fewest unpopped kernels?

Reading Comprehension

Consider the following statement:

"One thousand acres of rainforest are destroyed daily. As the rainforest is destroyed, hundreds of unidentified species are destroyed. We must stop rainforest destruction immediately to preserve biodiversity. It is wrong to destroy rainforests."

Which of the following questions would help evaluate the scientific accuracy of this statement?

a. How has the United States contributed to rainforest direction?
b. Why is it important to preserve the biodiversity housed in the world's rainforests?
c. Who is destroying the rainforest?
d. Where was the statement published?

Definition

A genus is composed of a number of related

a. Kingdoms
b. Phyla
c. Orders
d. Species

Verbal/Visual Extended

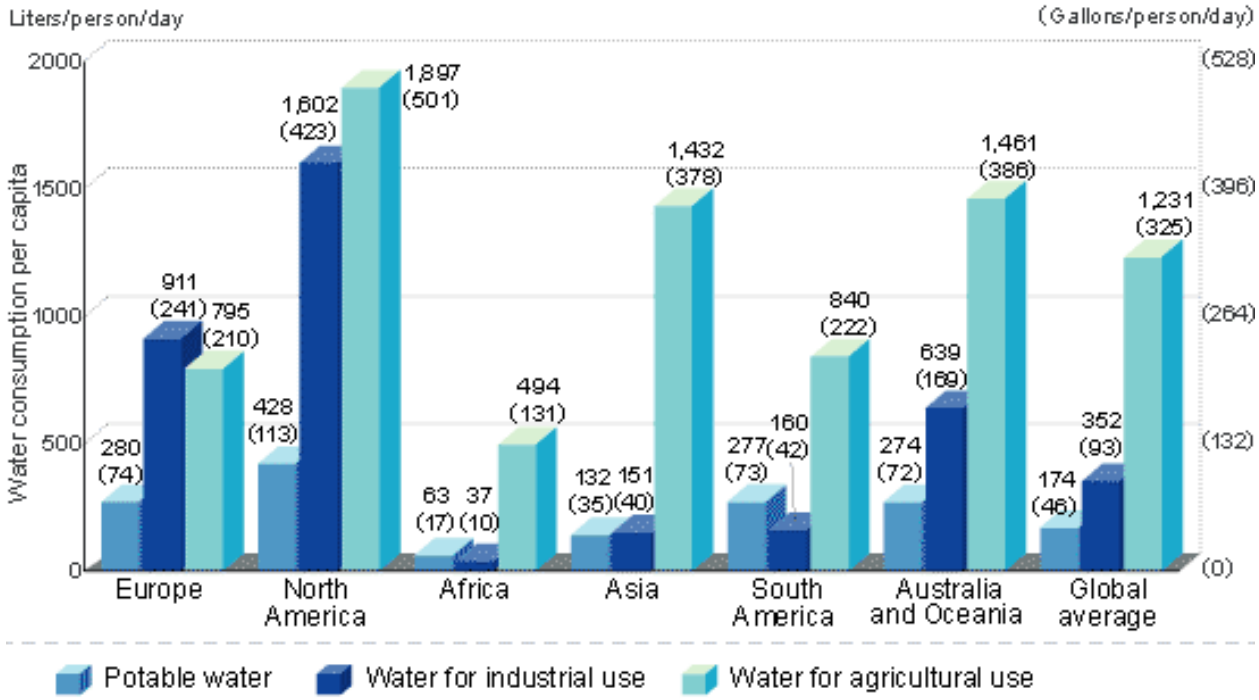
Creating Explanations

The year is 1925. You are the second speaker at the International Paleontology and Geology Conference in Silver Lakes, North Dakota. You have just listened to a two hour lecture on the impossibility of Pangaea given by Professor Dippidydo. It is now your turn to try to convince the audience of 2000 scientists that Alfred Wegener's evidence supports the existence of the supercontinent Pangaea 200 million years ago.

The lights dim. The crowd is silent. The spotlight turns on. What are you going to say?

Interpreting Data

Use the graph below to answer questions the following questions



The African continent has been plagued by periodic famines and human populations have faced starvation. What data does the chart present that helps understand why?

Which continent uses the most water per person to grow crops?

Basic Application

For the following chemicals:

1. Complete and balance the equation.
2. Identify the precipitate using solubility rules.
3. Write the net ionic equation for the following reaction.

Sodium carbonate plus silver (I) nitrate

Protocol Development

Using the experiment that we just did as a model, write up the directions for an experiment that will test your prediction of the effect that increasing the temperature of water would have on the rate of reaction between Alka-Seltzer and water.

Reading Comprehension

No examples were included in the submitted assessments.

Definition

What is an arthropod?