DEVELOPMENT OF THE *CHARACTERISTICS OF SCIENCE QUESTIONNAIRE (CSQ):* ASSESSING STUDENT KNOWLEDGE OF THE UTAH STATE SECONDARY SCIENCE CORE INTENDED LEARNING OUTCOME 6 ON THE NATURE OF SCIENCE

by

Bradford N. Talbert

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Instructional Psychology and Technology
Brigham Young University
September 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Bradford N. Talbert

This project report has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Richard R Sudweeks, Chair
Date	Nikki Hanegan
Date	Joseph Olsen
Date	David Williams
Date	Stephen Yanchar

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Bradford N. Talbert in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Richard R Sudweeks
Chair, Graduate Committee

Accepted for the Department

Andy S. Gibbons
Chair, Department of Instructional
Psychology and Technology

Accepted for the College

K. Richard Young
Dean, College of Education

ABSTRACT

DEVELOPMENT OF THE *CHARACTERISTICS OF SCIENCE QUESTIONNAIRE (CSQ):* ASSESSING STUDENT KNOWLEDGE OF THE UTAH STATE SECONDARY SCIENCE CORE INTENDED LEARNING OUTCOME 6 ON THE NATURE OF SCIENCE

Bradford N. Talbert
Department of Instructional Psychology and Technology
Master of Science

Teaching students about the nature of science is an important and necessary part of secondary science curricula. The Utah State Office of Education has provided specific guidelines called intended learning outcomes (ILOs) to teachers in the state. The ILOs are based on the national standards presented in the Project 2061 publications of the American Association for the Advancement of Science in 1990 and 1993. The ILOs are not tied to any one scientific discipline such as biology or chemistry, but are intended as global statements describing what scientists do and how scientific knowledge is gained.

ILO 6 prescribes that students be taught about the nature of scientific inquiry and the nature of the resulting knowledge claims. State education officials currently assess knowledge of the ILOs through items which are embedded in content-specific, multiple-choice items. This practice confounds knowledge of the nature of science with the content knowledge from a particular course. This project describes the development of an instrument to assess high school students' knowledge of the nature of science separate from their knowledge of any particular scientific discipline. The resulting questionnaire largely meets its intended goal, but still needs improvement. The current questionnaire has 24 items with $\alpha = .74$. Students participating in the pilot administration may have exhibited some apathy which may have affected the reliability estimate obtained from their responses. Factors such as this may help explain the low internal consistency reliability estimate of this questionnaire and will make the validity of the questionnaire difficult to demonstrate. In spite of this and other shortcomings, the questionnaire may still be useful to secondary science teachers to gain an understanding of their students' knowledge about the nature of science. Based on this knowledge, teachers may be able to design specific instruction to teach students correct knowledge about the nature of science. Since internal consistency reliability estimates are low, the validity of this questionnaire is tenuous. Therefore, caution should be exercised in making judgments about students that would affect measures of academic performance.

ACKNOWLEDGMENTS

I want to thank my wife Wendy. Without her encouragement I would never have started a Master's program. Without her encouragement I would never have finished one either. She has been patient with me on seventy times seven occasions while this project languished. She has been my help, support, and fan club for nineteen years. Thank you, Honey.

I also want to thank Dr. Richard Sudweeks. His advice and insights have been invaluable to me throughout this project. He has been patient when I wasn't working and enthusiastic when I was. I have found the breadth of his knowledge on various topics impressive and I have made unabashed use of his ideas.

To the members of my graduate committee: Thank you for your time, input, insights, and advice that have allowed me to bring this project to its current state.

And to Hugh Baird, Barb Gentry, Tom Erekson, Kevin King, and others who reviewed items and gave feedback: Thank you. Without your advice the items might have forever been mired in technical language beyond the reach of students.

TABLE OF CONTENTS

Page

APPENDIX

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter 1: Introduction

Science instruction in American high schools traditionally focuses on teaching subject-matter content in specific courses such as physics, chemistry, and biology. Students' acquisition of such content-based information is most frequently assessed through the use of paper-and-pencil tests consisting mainly of multiple-choice, short answer, and essay tasks. Students' understanding of the nature of science and the activities through which scientific knowledge is gained are usually more difficult to assess than content knowledge of a particular discipline. Assessing understanding of the nature of science often requires the researcher to rely upon students' individual self-reports (Netemeyer, Bearden, & Sharma, 2003). The difficulty of assessing students' understandings about the nature and practice of science is evident from the literature describing a variety of scales and their development (Aikenhead, Fleming, & Ryan, 1987; Chen, 2006; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Munby, 1997; Rubba & Andersen, 1978; Tedman & Keeves, 2001). The varied attempts highlight scholars' interest in assessing the beliefs about the nature and practice of science and the difficulties encountered in attempting to do so.

Science teachers are faced with the task of not only teaching subject-matter, but also with guiding their students into an appropriate understanding of the nature and practice of science. Teachers are also entrusted with helping their students develop mature and responsible attitudes about the proper application of scientific reasoning in daily life. To help teachers achieve these goals, the core curriculum in science developed under the auspices of the Utah State Office of Education (USOE) provides teachers with specific standards, objectives, and indicators for each grade level and subject.

Furthermore the Utah State Office of Education has provided intended learning outcomes (ILOs) that “describe the goals for science skills and attitudes” (USOE, 2002; p. 4) that students should acquire as a result of science education. The committee and persons responsible for the development of the core curricula and ILOs relied heavily on the publications of the American Association for the Advancement of Science (AAAS), *Science for all Americans* (1990) and *Benchmarks for Science Literacy: Project 2061* (1993).

The Utah state science ILOs apply to all elementary and secondary science content areas taught in Utah. The six secondary science ILOs specify that students should be able to do the following:

1. Use science process and thinking skills.
2. Manifest scientific attitudes and interests.
3. Demonstrate understanding of science concepts, principles, and systems.
4. Communicate effectively using science language and reasoning.
5. Demonstrate awareness of social and historical aspects of science.
6. Demonstrate understanding of the nature of science.

Each of these six ILOs is further defined by statements of specific knowledge, skills, attitudes, and understandings which students should manifest as evidence that they have attained the desired learning outcomes. End-of-level, multiple-choice tests are administered to all Utah high school students enrolled in physics, chemistry, biology, and earth systems science classes each school year. The items on these exams are aligned to both the content-specific standards and objectives in the core curriculum and to a specific statement within an ILO. For example, an item aligned to Standard I: Objective 2 in the physics core would simultaneously be aligned to ILO 1: Statement D. In this example, students would be asked to respond to the appropriate selection and use of a technological instrument (ILO 1: Statement D) to analyze the motion of an object in terms of velocity, time, and acceleration (Standard I: Objective 2). My personal experience as a high school science teacher has led me to conclude that using the described item format makes it nearly impossible to directly assess some of the stated objectives of the ILOs, particularly ILO 6.

ILO 6 of the Utah science core curriculum summarizes the various aspects and characteristics of the nature of science (NOS) presented in *Science for all Americans* (1990) and *Benchmarks for Science Literacy: Project 2061* (1993) presented in AAAS (1990, 1993). The aspects of the NOS as presented in ILO 6 essentially function as definitions of the various dimensions of the NOS. The Utah Core Curriculum in Science (2002) lists the individual aspects of the nature of science using nine statements denoted as *a*, *b*, *c,* etc. That same format will be used throughout this report. According to ILO 6, by the end of high school, students should understand the following ideas about the process of doing science and the characteristics of the resulting knowledge:

A. Science is a way of knowing that is used by many people, not just scientists.
B. Understand that science investigations use a variety of methods and do not always use the same set of procedures; understand that there is not just one "scientific method."
C. Science findings are based on evidence.
D. Understand that science conclusions are tentative and therefore never final. Understandings based upon these conclusions are subject to revision in light

of new evidence.

E. Understand that scientific conclusions are based on the assumption that natural laws operate today as they did in the past and that they will continue to do so in the future.

F. Understand the use of the term "theory" in science, and that the scientific community validates each theory before it is accepted. If new evidence is discovered that the theory does not accommodate, the theory is generally modified in light of this new evidence.

G. Understand that various disciplines of science are interrelated and share common rules of evidence to explain phenomena in the natural world.

H. Understand that scientific inquiry is characterized by a common set of values that include logical thinking, precision, open-mindedness, objectivity, skepticism, replicability of results, and honest and ethical reporting of findings. These values function as criteria in distinguishing between science and non-science.

I. Understand that science and technology may raise ethical issues for which science, by itself, does not provide solutions.

Statement of Purpose

The purpose of this project was to construct and refine a questionnaire for use in assessing students' understanding about the nature of science as described in ILO 6 of the Utah State Science Core Curriculum in Science. My intent was that the assessment scales resulting from this project would be useful to secondary science teachers in Utah as they assist their students in developing correct and healthy views about the NOS. A greater understanding by teachers of their students' views about the nature of science as an inquiry process and the nature of scientific knowledge should hopefully lead to the improvement of science instruction in the state. At the very least, I plan to use the resulting scales at my high school to guide teachers' efforts in this respect. If teachers want to teach students correct concepts about the nature of science they must first have some idea of what students already understand about the nature of science.

My hope is that by assessing their students' understanding of the NOS teachers will gain a greater understanding of their students' views of the nature of science and that a door will be opened for educators to discuss their own views about the nature of science. Ideally, this dialogue would lead to a greater focus and improved methods of imparting an appreciation for and an understanding of the nature of science to our students.

Research Questions

This project focused on answering the following questions:

1. Which items, if any, in draft versions of the instrument should be retained, revised, or deleted?
2. What is the dimensional structure of the set of nature of science items?
3. What is the internal consistency reliability of the resulting scales?
4. How well do the response categories function for each scale, and which, if any, should be collapsed or expanded?
5. What evidence supports the content validity of the scales?
6. What is the potential usefulness of the resulting NOS scales for science teachers to improve their instruction?

Chapter 2: Review of Literature

Throughout the years, various scales designed to assess students' understandings and views of the nature of science have been developed by scholars in the field of science education. One of the first challenges faced by researchers is to adequately and completely describe science and the nature of science. As a result, existing NOS scales vary considerably in organization, question format, and the exact nature of scientific knowledge assessed. Before discussing assessment scales that are currently in use, a discussion of the existing research into the nature of science itself seems appropriate.

The Nature of Science

One of the difficulties of discussing and assessing the nature of science is the inherent complexity of the subject. Philosophers of science do not completely agree about what exactly is meant by the term *nature of science* (NOS) (McComas, Clough, & Almazroa, 1998). Research indicates that the ability of a teacher to help their students develop healthy beliefs about the NOS depends upon the teachers' own understanding of the NOS (Lederman, 1992; McComas et al., 1998). McComas et al. (1998) imply that the lack of consensus among scholars raises the likelihood that classroom teachers will not receive in-depth instruction into the nature of science during their teacher education experience. Personal experience leads me to believe that if direct instruction is not given to pre-service teachers then the demands of daily classroom instruction leave little time for teachers to engage in personal research into the NOS.

The lack of complete consensus about a definition of the nature of science does not prevent curriculum specialists from identifying aspects of the NOS that teachers should be able to teach students. The indicators of ILO 6 (USOE, 2002) closely reflect the objectives summarized by McComas et al. (1998). The alignment of these two descriptions of the NOS as related to secondary education is demonstrated in Table 1. The first column of Table 1 summarizes common characteristics of the nature of science compiled from eight international sources (McComas et al., 1998). The second column restates the indicators of ILO 6. The entries in the second column are ordered so that parallel ideas in ILO 6 are aligned with corresponding ideas from the eight international sources. For example, statements 5 and 9 correspond to indicator F of ILO 6. Where no parallel statement exists *none* has been placed in the cell. Notice that ILO 6 does not contain statements expressing the role of creativity as scientists develop new insights, and novel hypotheses, methods, theories and explanations. Neither does ILO 6 describe the historical and social embeddedness of scientific inquiry; that science is both evolutionary and sometimes revolutionary. Finally, ILO 6 states that scientists assume that natural laws operate the same yesterday, today, and in the future. This idea is not paralleled in the list of NOS characteristics from the eight international sources. Even though scholars do not agree completely on how the NOS should be defined, the ideas stated in ILO 6 represent an attempt by the Utah State Office of Education to define and delimit the NOS domain sufficiently that it can be taught in Utah schools and assessed. The contents of Table 1 reveal that most, but not all, of the NOS as defined in ILO 6 of the Utah Science Core Curriculum is grounded in the work of previous scholars both from within the United States as well as from international sources. Hence, the contents of ILO 6 will be used as the working definition of NOS in this project.

Table 1

A Comparison of Utah Science ILO 6 with a Consensus of Science Objectives

Consensus objectives from international science standards documents[a]	Utah science ILO 6
1. Scientific knowledge, while durable, has a tentative character.	D. Understand that science conclusions are tentative and therefore never final. Understandings based upon these conclusions are subject to revision in light of new evidence.
2. Scientific knowledge relies heavily, but not entirely, on observation, experimental evidence, rational arguments, and skepticism.	C. Science findings are based on evidence.
3. There is no one way to do science (i.e. there is no universal step-by-step scientific method.)	B. Understand that science investigations use a variety of methods and do not always use the same set of procedures; understand that there is not just one "scientific method."
4. Science is an attempt to explain natural phenomena.	G. Understand that various disciplines of science are interrelated and share common rules of evidence to explain phenomena in the natural world.
5. Laws and theories serve different roles in science; therefore students should note that theories do not become laws even with additional evidence. 9. Observations are theory-laden.	F. Understand the use of the term *theory* in science, and that the scientific community validates each theory before it is accepted. If new evidence is discovered that the theory does not accommodate, the theory is generally modified in light of this new evidence.
6. People from all cultures contribute to science.	A. Science is a way of knowing that is used by many people, not just scientists.

(Table continues)

Table 1 (continued)

A Comparison of Utah Science ILO 6 with a Consensus of Science Objectives

Consensus objectives from international science standards documents[a]	Utah science ILO 6
7. New knowledge must be reported clearly and openly. 8. Scientists require accurate record keeping, peer review, and replicability.	H. Understand that scientific inquiry is characterized by a common set of values that include logical thinking, precision, open-mindedness, objectivity, skepticism, replicability of results, and honest and ethical reporting of findings. These values function as criteria in distinguishing between science and non-science.
10. Scientists are creative. 11. The history of science reveals both an evolutionary and revolutionary character. 12. Science is part of social and cultural traditions. 13. Science and technology impact each other. 14. Scientific ideas are affected by their social and historical milieu.	None
None	E. Understand that scientific conclusions are based on the assumption that natural laws operate today as they did in the past and that they will continue to do so in the future. I. Understand that science and technology may raise ethical issues for which science, by itself, does not provide solutions.

[a] First column entries from McComas, Clough, and Almazroa (1998). The role and character of the nature of science in science education. In W.F. McComas (Ed.), *The Nature of Science in Science Education: Rationales and Strategies* (p. 6). The Netherlands: Kluwer Academic Publishers.

ILO 6 of the Utah science core prescribes what teachers are expected to teach their students about the nature of science. This requirement necessitates that teachers themselves learn about NOS and how to teach it effectively. Teachers of science play a vital role in providing students with healthy beliefs about the functions, processes, and limits of science.

The Twofold Nature of Scientific Knowledge

The word *science* refers both to the process of conducting scientific inquiries and the knowledge gained from such inquiries; it refers to a process and the product of that process (Good, 2000; Hogan, 2000; McComas et al., 1998). The indicators in ILO 6 reflect these two different perspectives of the NOS in this manner: (a) characteristics of the inquiry process that typify scientific investigations (indicators a, b, g, h, and possibly f) and (b) the characteristics of the kinds of knowledge claims that typically result from scientific inquiries (indicators c, d, e, and i). Understanding of the practice of science is critical to scientific literacy (Hogan, 2000) because people who understand how scientific knowledge is produced have a better foundation for critically and thoughtfully evaluating the resulting knowledge claims.

Scientific knowledge is gained through empirical observations. These observations are supposed to be objective in nature and potentially replicable (Lederman, 1992). Scientists value skepticism, objectivity, creativity, and honesty in their work (McComas et al., 1998). These values describe or encompass the process of doing science and help to distinguish science from pseudoscience. Knowledge gained through inquiries that adhere to these scientific values is generally given greater credibility. The process of doing science, therefore, leads to new knowledge through empirical tests conducted as objectively as possible. The results of scientific tests are skeptically reviewed by a community of scientists who share common scientific values in their work.

Knowledge claims gained as a result of implementing scientific inquiry are often thought of as science. Knowledge gained through the practice of science is likely to be accepted by people because of the processes by which it was obtained. Such knowledge should be based on empirical evidence, produced by creativity, reported honestly, and subjected to skeptical review before it is accepted by the scientific community (McComas, 1998). Scientists generally strive to be objective in their work, but it must be noted that true objectivity is nearly impossible to attain because scientific inquiry is embedded in the social and historical milieu of the researcher. Non-scientists tacitly expect adherence by scientists to values such as skepticism, objectivity, creativity, and honesty when they accept scientific claims.

Knowledge of both the procesess of science as well as the resulting products is necessary for the layperson to be able to critically evaluate scientific claims (Hogan, 2000). Such knowledge is invaluable for making informed, educated decisions about the scientific or non-scientific nature of claims which purport to be based on scientific analysis (Scharmann & Smith, 2001; Smith & Scharmann, 1999).

Describing the Nature of Science

McComas & Olson (1998) describe an extensive effort on their part to identify the elements which best describe the NOS at a level that is appropriate for science learning contexts. Their stated goal was to “produce a definition of the content of the nature of science (NOS) useful in informing science teaching and learning” (p. 41). To this end they performed a qualitative survey of the science education standards of several countries and identified those characteristics of the NOS that appeared in each document. They classified the NOS statements into four broad categories, namely (a) philosophical, (b) sociological, (c) historical, and (d) psychological. McComas et al. (1998) describe the NOS this way:

> The nature of science is a fertile hybrid arena which blends aspects of various social studies of science including the history, sociology, and philosophy of science combined with research from the cognitive sciences such as psychology, into a rich description of what science is, how it works, how scientists operate as a social group and how society itself both directs and reacts to scientific endeavors. (p. 4)

No single widely accepted definition of the nature of science is given in the literature. Instead the NOS is described in terms of various traits which a scientifically literate person demonstrates or articulates about the way in which science is practiced in the pursuit of new knowledge (Lederman et al., 2002; Rubba & Andersen, 1978). Rubba and Andersen (1978) describe scientific knowledge as (a) amoral, (b) creative, (c) developmental, (d) parsimonious, (e) testable, and (f) unified. Lederman et al. (2002) describe the NOS by describing the different but cooperative roles played by (a) observation and inference, (b) theories and laws, (c) creativity and imagination, (d) subjectivity and theory-ladenness of scientific knowledge, and (e) the social and cultural embeddedness of scientific knowledge.

McComas et al. (1998) summarized nature of science objectives from eight international science standards documents. Their summary closely matches the descriptions presented by other researchers (Lederman et al., 1998, 2002; Aikenhead et al., 1987; Clough, 1998; Lederman, 1992; Rubba & Andersen, 1978; and others) as outlined in Table 1.

Aspects of the Nature of Science

Many researchers in science education have attempted to describe and define the nature of science. These researchers do not agree completely about the specifics of what constitutes the nature of science and exactly what secondary science students need to know about the processes and products of scientific investigations. However, from a comparison of research done by Rubba and Andersen (1978), Lederman (1992), Niaz (2001), Scharmann and Smith (2001), and Smith and Scharmann (1999), several parallels can be found. A comparison of the descriptors suggested by these researchers with the indicators listed in ILO 6 of the Utah secondary science core document (USOE, 2002) shows good agreement as demonstrated in Table 1. In the following paragraphs, I present summaries describing common essential characteristics of the nature of science found in the literature.

Value laden. Scientists employ values in their work. These values include (a) logical thinking, (b) skepticism, (c) open-mindedness, (d) objectivity, (e) replicability, (f) accuracy, and (g) honesty in reporting (AAAS, 1990, McComas et al., 1998, Smith & Scharmann, 1999). Scientific inquiry arises from the curiosity of scientists to describe some feature of the natural world. Assumptions and values held by scientists naturally infuse their work. However, when interpreting results, scientists attempt to do so in an objective manner. As in other pursuits, interpretation of scientific observations is invariably colored by personal experience and philosophy and often by the researcher’s goals and preconceptions. A certain amount of subjectivity on the part of researchers is inevitable (Lederman et al., 2002). Although they may not always succeed in doing so, scientists seek to judge their own work and the work of others in a detached manner, attempting to be objective in their views and open to points of view other than their own (AAAS, 1990). However, different scientists will often interpret the same results in different ways (Niaz, 2001). This leads to ongoing discourse and debate inside the larger scientific community which gives science much of its richness and diversity. Values held by scientists include the importance of being honest in fully disclosing unfavorable as well as favorable results (McComas et al., 1998).

Empirical. Scientific inquiries are based on sensory data and phenomena which can be observed. Scientific findings are based on evidence and analysis (AAAS, 1990, 1993; Lederman et al., 2002; McComas, 1998; McComas et al., 1998). Consistency and replicability of results obtained over time and by several researchers lends validity and credibility to scientific findings (Smith & Scharmann, 1999). Scientific inquiry arises from the curiosity of scientists to describe some observation in the natural world. Productive scientific work requires creativity to devise tests and propose theories and explanations to describe observations.

Developmental. Scientific research often leads to new explanations and theories. New knowledge gained by scientific inquiry is necessarily tentative and fallible (Lederman et al., 2002; McComas, 1998). Since scientific knowledge claims are based on observations, they are often the product of inductive reasoning. Scientific claims are based on a preponderance of the evidence available at any given time. Therefore, as new inquiries are performed and new evidence is obtained, scientific claims often need to be reformulated in light of the new information (AAAS, 1990; McComas et al., 1998). The accumulation of knowledge over time is what lends scientific work its tentative nature. Even the most strongly held beliefs are susceptible to revision in light of new or contradictory evidence. This is not to say that such changes occur at the first sign of contradiction, but only after careful review by the larger scientific community. The development of new ideas, including gathering convincing evidence, may take considerable time and may go through several iterations before being accepted by the scientific community.

Explanatory. Science seeks to explain natural phenomena (AAAS, 1990, 1993, McComas, 1998; McComas et al., 1998). Scientific knowledge is a vast network of interrelated concepts, theories, and laws that together have predictive (AAAS, 1990) and explanatory power. One of the purposes of scientific inquiry is to generate powerful explanations with great predictive power of natural phenomena. Such explanations rarely arise from one person working in isolation. In fact, much scientific discovery is the result of collaboration between members of a larger scientific community. Observations and explanations are put forth for review and critique by other scientists (McComas, 1998). This process builds community and collegiality if not consensus among scientists. Through such efforts within a larger community, interpretations of scientific results can gain their greatest explanatory and predictive power. Science seeks parsimonious explanations for observed phenomena (Rubba & Anderson, 1978). Simple explanations with the greatest explanatory and predictive power are preferred in contrast to unnecessarily complex explanations.

Open to public scrutiny. Science is intertwined with the social, political, and historical context in which it is performed and interpreted (AAAS, 1990; Lederman et al., 2002; McComas et al., 1998). Scientific research is subject to review, interpretation, reinterpretation, criticism, and acceptance or rejection by other scientists, as well as by politicians and the public at large. All are potential consumers of research findings and conclusions. While scientists might hope that research would be free from any influence except the search for pure scientific knowledge, they would be foolish to completely ignore the opinions of financial underwriters and the public to whom scientists must publish the results of their work. Scientists must not only report their work in a manner clear enough for other scientists to understand, but scientists must also report their work clearly enough and in sufficient detail for other informed laypersons to understand. Those who fail to do so risk having their work misrepresented at best and discredited at worst. Scientists do not work in a vacuum, but rather in a social framework that includes colleagues, peers, politicians, and the lay public.

ILO 6 of the Utah State Science Core

The Intended Learning Outcomes of the Utah State Secondary Science core (USOE, 2002) state that the main goal of science instruction in Utah schools is for students to “value and use science as a process of obtaining knowledge based upon observable evidence” (p. 4). ILO 6 consists of nine definitions of different aspects of the nature of science. A greater understanding of the nature of scientific processes should reasonably lead to a greater understanding of the nature of the scientific knowledge claims made as a result.

The following paragraphs discuss each of the definitions under ILO 6 and demonstrate how each blends with the descriptions of the NOS which were shown in Table 1. The USOE (2002) document lists the indicators as *a, b, c* etc. The same format is used here.

Indicator A: Science is a way of knowing that is used by many people, not just scientists. People may claim to know something through scientific evidence, social interactions, authority figures, religious teachings and experiences, and cultural norms. This indicator speaks of science as only one way of knowing. It is connected to the empirical and explanatory aspects of the NOS. In an ideal scientifically literate society, all people would believe that the world is understandable and they would be capable of evaluating empirical evidence in order to decide the degree of truth in claims of scientific knowledge (AAAS, 1990, 1993). People should be able to examine and interpret multiple representations of information (e. g. graphical, tabular, mathematical) in order to develop informed opinions.

Indicator B: Understand that science investigations use a variety of methods and do not always use the same set of procedures; understand that there is not just one "scientific method." Scientific work is creative and value-laden (McComas, 1998; McComas et al., 1998). In order to devise methods to test theories and hypotheses, scientists must be creative, not locked into any single method that fits all types of inquiry (AAAS, 1990, 1993; Lederman et al., 2002; Rubba & Anderson, 1978; Smith & Scharmann, 1999). The values, beliefs, and background of the researcher all influence the methodological choices which he or she makes. Indeed, decisions about which questions to investigate and which hypotheses to propose are inextricably linked to the scientist's previous experiences, expectations, and personal values (Lederman et al., 2002; McComas et al., 1998). In addition, the researcher's interpretations of the data collected and conclusions drawn from the data are influenced by the researcher's perspective and values.

Indicator C: Science findings are based on evidence. Scientific knowledge depends upon the collection of observations that are capable of being replicated by other observers working independently (AAAS, 1990; Lederman et al., 2002; McComas et al., 1998; Rubba and Anderson, 1978; Smith & Scharmann, 1999). To collect data scientists may rely on their own senses, instruments which enhance the human senses, or instruments which detect phenomena unobservable by human senses (AAAS, 1990). Thought experiments may provide guidance and good theory is necessary, however, scientific claims are not accepted in the absence of empirical evidence.

Indicator D: Understand that science conclusions are tentative and therefore never final. Understandings based upon these conclusions are subject to revision in light of new evidence. This indicator ties into many aspects of the NOS; namely, science is empirical, developmental, and explanatory (Rubba & Anderson, 1978). Because of its dependence upon empirical data, scientific knowledge must be tentative (Lederman et al., 2002, McComas, 1998, Smith & Scharmann, 1999). To believe that all possible observations have been made or that all possible data have been collected defies logic. If further studies fail to replicate results or if the results completely refute previous claims, then both the old and the new knowledge claims are subject to further scrutiny and debate (AAAS, 1990). Convincing evidence is merely that, convincing. It does not establish finality (McComas, 1998) or exclude the possibility of the future invalidation of current theories.

Indicator E: Understand that scientific conclusions are based on the assumption that natural laws operate today as they did in the past and that they will continue to do so *in the future.* A necessary assumption of science is that the laws underlying and underpinning the observations which scientists make have and always will operate the same (AAAS, 1990). What does change is the current understanding of those laws. In *Science for All Americans* the AAAS (2002) claims that "science also assumes that the universe is…a vast single system in which the basic rules are everywhere the same" (p. 2). What scientists do not assume, however, is that our current knowledge or explanation of those laws (basic rules) are complete.

Indicator F: Understand the use of the term "theory" in science, and that the scientific community validates each theory before it is accepted. If new evidence is discovered that the theory does not accommodate, the theory is generally modified in light of this new evidence. Theories are created by scientists to explain data collected in experiments and nonexperimental investigations. The data collected by scientists allow people to glimpse evidence of the operation of underlying natural laws. Theories share a different and vital role from laws, but they do not mature into them (McComas, 1998). Lederman et al. (2002) explain the difference between theories and laws in this way, "In general, laws are descriptive statements of the relationships among observable phenomena. . . . Theories, by contrast, are inferred explanations for observable phenomena or regularities in those phenomena" (p. 500). Theories are frequently debated in public forums such as scholarly journals. Theories are often revised to accommodate new data or to exclude discredited ideas. All other considerations being equal, scientists generally prefer the simplest theory to explain extant data.

Indicator G: Understand that various disciplines of science are interrelated and share common rules of evidence to explain phenomena in the natural world. Science is often thought of as being divided into various disciplines such as chemistry, biology, physics, geology, zoology, etc. The division of scientific knowledge into disciplines provides a structural framework for organizing research and research findings. These divisions may not be reflected in nature and may occasionally make communication between scientists more difficult. Because nature is not divided into disciplines, considerable overlap exists between the various sciences (AAAS, 1990). Biology blends into chemistry, physics, paleontology, and psychology. The various disciplines are all equally scientific and together comprise the scientific endeavor. Since the various scientific disciplines share common rules of evidence, scientists working in a collaborative manner can greatly enrich the body of cumulative scientific knowledge by approaching the formation of explanations and theories from multiple points of view (Rubba & Anderson, 1978).

Indicator H: Understand that scientific inquiry is characterized by a common set of values that include logical thinking, precision, open-mindedness, objectivity, skepticism, replicability of results, and honest and ethical reporting of findings. These values function as criteria in distinguishing between science and non-science. Scientific work and scientific knowledge are value-laden (AAAS, 1990; McComas et al, 1998; Smith & Scharmann, 1999). The values described in this indicator function as criteria that scientists use to judge their own work as well as the research of their peers. When reviewing the work of their peers, scientists expect that their peers adhere to the values listed. Like any other activity, the rigorousness of adherence to these values varies from person to person. However, adherence to these values can be used as a litmus test to distinguish scientific claims from non-scientific claims (Smith & Scharmann, 1999).

Indicator I: Understand that science and technology may raise ethical issues for which science, by itself, does not provide solutions. Science is limited to exploring questions that can be answered empirically. Questions of an ethical, moral, or religious nature are also important, but cannot be answered solely by empirical evidence (AAAS, 1990; Rubba & Anderson, 1978). Science can expound upon what is, or at least what seems to be, based upon the best possible explanation of data. Scientific inquiry cannot reveal how that knowledge is best put to use. As a result, science may be unable to answer all the questions that may arise from the application of scientific knowledge.

Aspects of the NOS not included in ILO 6. Aspects of the NOS not adequately represented in ILO 6 are (a) the role of public dissemination and public scrutiny in scientific work (b) the inherent subjectivity of scientists when performing scientific inquiry, and (c) the creative human element of scientific inquiry. Of necessity, science is a public endeavor. Scientists are expected to submit their work for peer review and public dissemination. This review is usually required to obtain some level of acceptance of the work being reported and often to obtain or retain project funding. Science is also subjective in nature. Hypotheses, methods, analyses, and interpretations of results are all subject to the particular point of view of the researcher. Since scientific inquiry is a human endeavor, personal creativity plays a vital role as scientists formulate hypotheses, devise experimental procedures, and propose insightful interpretations of data.

Assessing the Nature of Science

A preliminary search for existing instruments revealed several scales designed with the intent of assessing students’ understanding of the nature of science and/or attitudes towards science (see Aikenhead et al., 1987; Rubba & Andersen, 1978 for examples). However, a careful reading reveals many items and underlying constructs that are not aligned to the Utah state ILOs. While these instruments may provide a fruitful source of ideas and may be useful for other purposes, they do not align sufficiently with ILO 6 to adequately assess the NOS as defined in ILO 6 of the Utah Core Curriculum in Science. I believe that assessment of students understanding of the nature of science as outlined by ILO 6 could be best accomplished by writing new items that are more closely aligned to the Utah core than the items of existing instruments. Additionally, Lederman, Wade, and Bell (1998) point out that most of the instruments available address only certain aspects of the nature of science and often inappropriately confuse the nature of science with attitude towards science. Lederman, Wade, and Bell (1998) surveyed extant NOS measurement scales. Many were found to lack validity as NOS measurement devices. Here is a brief summary of some of the most commonly used NOS measurement scales often considered by researchers to yield results with the greatest validity.

Test on Understanding Science (TOUS)

This test was developed by Cooley and Klopfer in 1961. The TOUS was very widely used by researchers to assess knowledge and beliefs about the nature of science. The TOUS consisted of 60 multiple-choice items with four alternatives each. It was further divided into three subscales designed to assess (a) understanding about the scientific enterprise, (b) the scientist, and (c) the methods and aims of science. Many of the items however, related more to a student's conception of the profession of scientist than the nature of science as it is currently described in the literature. Many of the items may have elicited an emotional response about the propriety of a particular policy or practice instead of assessing a student's actual knowledge of the NOS. Lederman et al.

(1998) point out that the TOUS may not be appropriate as the sole instrument to measure students' understanding of the NOS.

Views on Science-Technology-Society (VOSTS)

VOSTS was developed and validated for grade 11 and 12 students by Aikenhead et al. (1987). One of the underlying assumptions in the development of this instrument was that researchers and students do not necessarily attach the same meanings to a particular concept or word. Therefore, written student responses to open-ended questions about science, technology, and society were analyzed and Likert-type items were generated from the students' responses. While many statements in the VOSTS instrument may be useful for ideas to develop items from, the VOSTS instrument encompasses far more than the current project, such as (a) defining science, technology and the relationship between them, (b) defining research and development, (c) asking opinions on how government money should be disbursed, (d) probing opinions about what entity should control research, and (e) asking about ethical and religious implications.

Nature of Scientific Knowledge Scale (NSKS)

Developed by Rubba and Anderson in 1978, the NSKS consists of 48 Likert-type items with five choices. The model of scientific knowledge which underlies the development of this instrument is fairly basic and aligns reasonably well with the NOS as presented in Table 1. Rubba & Andersen (1978) developed a six-factor model of the nature of science in which they describe scientific knowledge as (a) amoral, (b) creative, (c) developmental, (d) parsimonious, (e) testable, and (f) unified. While these six factors are represented in Table 1, they are not a sufficient description of the NOS. The NSKS includes eight items for each of these factors. Four of the eight items for each factor are positively worded and four are negatively worded. Therein lays the potential problem as I see it. Many of the negatively worded items are the positive items with simple negating language inserted (Rubba & Andersen, 1978). Two items associated in this way will have correlated error variance in the responses. This may have the effect of inflating the internal consistency reliability estimates obtained for the NSKS since students responding may just pick the opposite response from an opposite item answered previously.

Views of Nature of Science Questionnaire (VNOS)

This assessment of students' views about the nature of science was developed by Lederman et al. (2002) in response to their assertion that the best way to assess students' understanding of the NOS is through open-ended interview-type questions. Consequently, the VNOS consists of ten questions presented in an open response format to which students give written responses. Chen (2006) points out that it may be difficult for students to adequately articulate their responses to the prompts in this instrument in a 40 to 60 minute time frame. Many teachers are unwilling or unable to invest one or two class periods to the administration of a survey. Likewise, interpretation of responses is likely to be very time consuming for teachers, especially in light of the advice given by Lederman et al. (2002) that a large proportion of students should be interviewed after taking the questionnaire to validate responses. While the practice of interviewing is practical and necessary to the researcher, it is impractical to expect classroom teachers to do so extensively.

Reliability and Validity

Much attention is given in the literature to the question of measurement validity. Generally, establishing validity involves providing evidence that a measurement instrument does indeed measure what it purports to measure. The literature promotes the use of more than one type of evidence to establish validity. Anderson and Bourke (2000) indicate that for self report affective assessments, validity can be established through two methods: evidence of content validity and evidence of construct validity.

Reliability generally refers to the amount of random error present in measurements. Psychometrically, reliability depends on the ratio of true variance to the total variance of the observed scores from a measurement. Practically, reliability indicates the amount of confidence that can be placed in a particular assessment to elicit similar behavior from respondents over multiple administrations of the assessment. Reliability is a numerical summary of the consistency of two or more measures of the same trait obtained from the same respondents.

When the items in a scale correlate highly with each other, reliability will be maximized. However, validity is increased when the items of an assessment correlate highly with some external variable. Dawis (2000) states, “In a scale with high inter-item correlations, the items all measure the same thing, which in turn minimizes the content domain coverage of the scale and diminishes the scale’s chance of correlating with an external variable” (p. 86). Loevinger (1954) referred to this trade-off as the *attenuation paradox*. While reliability is a necessary precursor to validity, to make both reliability and validity optimal, neither can be maximized (Dawis, 2000).

Summary

The nature of science is a rich and complex topic. Scholars do not agree completely about how to define it or how to best assess students’ understanding of it. Nevertheless, many scholars, curriculum theorists, and philosophers of science agree on a set of core characteristics that should be included in any attempt to teach high school students to understand the nature of scientific inquiry and knowledge. These central characteristics include the following:

1. Science is a way of seeking and acquiring knowledge based on observable evidence that can be replicated.
2. Science is a human endeavor aimed at helping us to understand and explain natural phenomena.
3. The process of deciding what questions or issues to investigate and what methods to use to collect evidence is theory-laden. Likewise, the process of analyzing and interpreting the resulting observations is also theory-laden. Hence, scientific inquiry can never be completely objective.
4. Since scientific research is an inductive, developmental process that builds upon itself as more and more evidence is acquired, scientific knowledge is always tentative and subject to revision as new evidence is accumulated.
5. The inquiry methods employed in a particular investigation as well as the findings obtained and the manner in which they are interpreted should be publicly disclosed and subjected to scrutiny by other researchers and by the general public.

ILO 6 in the current Utah Core Curriculum in Science aligns well with these aspects with some differences. First, ILO 6 does not include statements adequately expressing the social and historical embeddedness of scientific inquiry. Second, ILO 6 does not address the inability of scientists to be completely objective. Their work is always colored by their own epistemology and background. Third, ILO 6 fails to mention the creative element required in scientific research. Lastly, ILO 6 makes specific mention of the assumption that natural laws will always operate in the same manner through past, present, and future. This may very well be a necessary assumption underlying much scientific work; however, this assumption is not mentioned in the literature and is therefore specific to the Utah Core Curriculum in Science.

Assessing what students believe about the NOS is a difficult task. Lederman et al. (1998) criticized the idea of using traditional paper-and-pencil questionnaires to assess students' understanding of the NOS. They contend that students do not necessarily interpret the statements on these instruments in the same way as the researcher who wrote them. Munby (1982) referred to this concern as one of the primary weaknesses of validity estimates obtained by the so-called panel of experts. Aikenhead et al. (1987) deliberately attempted to avoid the problem of student misinterpretation of items in the development of the *Views on Science-Technology-Society* scale. To avoid item interpretation errors by students, many researchers advocate the use of open-ended constructed-response assessments of NOS as well as conducting extensive interviews with students to ascertain what students meant in their written responses (Lederman et al., 1998, 2002; Good, 2000).

For the researcher, lengthy surveys and personal interviews may be best. For the classroom teacher, daily interaction with students provides many opportunities to probe what students know about the NOS and to develop specific instructional activities designed to assist students in developing healthy and correct views about NOS. For this type of daily interaction and adjustment, a baseline assessment of student knowledge may be gained through a questionnaire-type assessment. I believe that within the core characteristics of the NOS described in ILO 6 students possess correct and incorrect views about the nature of scientific inquiry and scientific knowledge claims. These views can be assessed by using declarative statements to which respondents indicate their level of agreement with the statements. The same assessment could be given at the end of an instructional unit or course to learn if activities and lessons met the intended goal of teaching the nature of scientific knowledge. This form of assessment and instruction assumes that the science teacher has a correct understanding of the NOS. However, this last assumption may prove to be the greatest obstacle to high school students gaining correct views about the NOS.

Chapter 3: Method

Items to assess secondary students' knowledge about the nature of science were written in two phases. In the first phase 45 items were written and tested. Based on reliability and Rasch analyses some of these items were revised, others were deleted, and new items were written to fill in the resulting gaps in the indicators of ILO 6. Subsequent analysis revealed that many of the items did not align as well with ILO 6 as had been intended. Because many of the items had undergone revisions to align them better with ILO 6 and because there were several new items a second pilot was administered to students. The second pilot involved testing and analyzing 38 items in a manner similar to the first pilot to obtain a more refined assessment.

Target Population

The *Characteristics of Science* (CSQ) measurement scale was designed for use in high school science classes in the state of Utah. Because the ILOs are global statements that span across content areas, students responding to the scale could be enrolled in any science class taught at any high school in Utah. The scale could be used at the beginning and end of each academic year to assess changes in students' beliefs that may have resulted from instruction.

Items in the scale were specifically targeted towards students in the ninth through twelfth grades. These students range in age from 14 to 18 years of age. Students in this age group generally have completed science classes in grades seven and eight. The ILOs apply to all secondary science classes from grade 7 to grade 12. It might be assumed that students have already had some exposure, either direct or indirect, to the goals specified in the ILOs. However, research by Abd-El- Khalick (2000), Lederman et al. (1998), and

Martin-Diaz (2006) among others highlights the frequent lack of training in the NOS received by science teachers. Therefore, there is little evidence to support the assumption that secondary students would have been specifically instructed in the NOS.

Item Development

While ideas for possible items were taken from existing scales, the items for the CSQ are original and were written specifically for this instrument. Each item is a declarative statement describing some aspect of the nature of scientific knowledge or scientific enquiry. Statements are based directly on the nine indicators of ILO 6 with items included to address aspects of the NOS not addressed in ILO 6 as described previously. Students were asked to respond to each statement by indicating their personal level of belief or agreement with the statement on a six-category response scale. The response options were 0 (*strongly agree*), 1 (*agree*), 2 (*slightly agree*), 3 (*slightly disagree*), 4 (*disagree*), and 5 (*strongly disagree*). Particular attention was given to the task of writing statements which would be understood by the target audience.

The goal of this project was to provide secondary science teachers in Utah with information about their students' knowledge and beliefs about the nature of science as outlined in ILO 6 of the Utah Core Curriculum in Science. The questionnaire resulting from this project may provide a basis for judging student conceptualizations of the NOS as correct or incorrect, and also provide a starting point for open dialogue among science teachers about how to improve instruction about the NOS in our classrooms. However, it is not intended as a vehicle for course grade determination.

Establishing Validity

Evidence of content validity can be demonstrated using a panel of experts to review items. Subject matter experts independently rate the items according to what they believe the items measure. Munby (1982, 1997) suggests that validity determined by a panel of experts is dependent upon the assumption "that the meanings test items have for judges are in some way equivalent to those held by the students who are to take the test" (p. 338). This assumption may be tenuous since meaning and understanding are largely based on education and experience.

Throughout this project the researcher could not expect that students had either the same education or experience as the judges. Evidence of content validity did not rely solely on the judgment of a panel of experts, but was also obtained by interviewing students prior to administering the pilot questionnaire to ascertain their interpretation of the items (Good, 2000; Lederman et al., 1998). Lederman et al. (1998) assert that interviewing is the only way to obtain valid and reliable knowledge about what students know of the nature of science. Through interviews, students' responses on items from the assessment can be corroborated. Interviews allow the researcher to verify students' understanding and interpretation of individual items. Interviews may reveal that students do not interpret an item as intended or that an item functions in an unintended but desirable manner. By interviewing students about their interpretations of the items and making appropriate adjustments prior to administering the questionnaire content validity was built into the CSQ.

Evidence of construct validity can be established from analysis of data collected from administration of the assessment. Since this project was to produce a scale reflecting several indicators from ILO 6, construct validity can be established by demonstrating that the items group together according to the indicators in ILO 6. This was accomplished using exploratory factor analysis (Gardner, 1995).

Providing adequate evidence of both content and construct validity helps establish that the scale measures what was intended. An instrument which would provide for valid inferences and conclusions was developed in two ways (Anderson and Bourke, 2000). First, great care was taken to write items which could be clearly classified as true or false by persons experienced and knowledgeable about the scientific enterprise. Second, prior to administration of the questionnaire, members of the target audience were interviewed to establish their interpretation and understanding of the items.

Review by experts. Items were submitted through several iterations to people knowledgeable about the nature of science and about the Utah Core Curriculum in Science. This panel of experts was comprised of science curriculum officials employed by the Utah State Department of Education, university professors, and secondary science teachers. Each individual was asked to review the items for their alignment to ILO 6, their representation of correct conceptualizations of the NOS, and their suitability for secondary students. Based on recommendations made by this group, items were revised until consensus was reached.

Review by representatives of the target audience. At several stages as items were developed, students enrolled in secondary science classes were asked to review the items and to comment on their readability and meaning. Students were encouraged to talk out loud about the items as they read and responded to them. Particular attention was paid to vocabulary and the meaning extracted by the students. Items were revised based on this feedback. Care was taken to select student reviewers representing different grade levels and general scholastic aptitude.

Pilot Testing

Once items had been reviewed and revised they were pilot tested with members of the target audience. In both pilot administrations of the CSQ, students came from the tenth through twelfth grades only. Pilot testing was performed in two stages. The first pilot, administered in the spring of 2006, provided information about item and response category functioning. The second pilot, administered in spring of 2007, provided similar information and allowed for final item revisions.

First pilot administration. In the spring of 2006 the initial draft version of the CSQ consisting of 45 items was pilot tested with 353 students in biology, chemistry, and physics classes in a Utah secondary school. Science classes at this high school are comprised of students in the tenth through twelfth grades. Groups consisted of students drawn from basic and advanced courses in the three subjects listed. The questionnaire asked students to respond by indicating the degree to which they agreed or disagreed with the proposition expressed in each statement using a six-point response continuum: 1 (*strongly agree*), 2 (*agree*), 3 (*slightly agree*), 4 (*slightly disagree*), 5 (*disagree*), and 6 (*strongly disagree*).

Data analysis. Data collected from the pilot questionnaire were analyzed to identify which items functioned best at assessing the NOS described in ILO 6. Before performing any statistical analysis of the questionnaire data, the response sheets were examined for any patterns indicating that a particular student might have been simply filling in bubbles without reading the items and mentally processing the propositions expressed. These students were identified by examining their response sheets as well as the person-fit statistics from the Winsteps software program. Any students found to have suspect response patterns or poor infit mean-square values were deleted from the questionnaire results. Questionnaire data were then analyzed using item-to-adjusted-total correlation coefficients and item-fit statistics obtained from Winsteps. Items with low item-total correlations and outfit mean-square values greater that 2.0 were examined and either modified or deleted. The effect of deleting items on the reliability and validity of the scale was also considered. Items were primarily rewritten instead of being deleted if at all possible.

Once items had been identified, deleted, or revised, gaps in the coverage of the indicators in ILO 6 were identified. New items were written with a greater attention to those aspects of the NOS which were deemed to be essential characteristics of a healthy conceptualization of the nature of science.

Second pilot administration. Revised and new items were piloted in the spring of 2007 with a new group of students from a population similar to the first pilot. The Winsteps program was again used to perform person analysis, item analysis, reliability analysis, and analysis of the response categories in a manner similar to the item analysis performed on the first pilot of the questionnaire. Factor analysis was used to help determine the underlying structure of the CSQ and the aspects of the NOS it describes.

Chapter 4: Results

To address the research questions of this project, item response data were analyzed utilizing the Rasch model with the Winsteps software program, reliability analysis, and factor analysis. Evidence of the items functioning as intended to assess the nature of science was necessary.

First Pilot Administration

The first version of the CSQ, consisting of 45 items, was administered to 353 secondary science students in the spring of 2006. Items for this first pilot were aligned with the nine indicators of ILO 6 mentioned previously. The Winsteps software, developed by M. Linacre in 1998, was used to analyze the students' responses using a combination of Rasch techniques and traditional item analysis. Item analysis statistics are presented in Table 2.

Item Retention, Revision, or Deletion

Item-total correlations. Traditional item analysis begins by examining the item-total correlation values. Generally correlation values .30 and above are desirable. A glance down the third column of Table 2 reveals the following items with low correlation values: Items 2, 3, 4, 10, 22, 27, 36, 37, 39, 41, 42, 43, and 45. These items were examined further for revision or deletion; the others will be retained. Special note should be taken of Items 2, 3, 22, 36, 39, 41, 42, which were negatively worded on the CSQ pilot. Before calculating the statistics presented in Table 2, these seven items were reverse scored. Therefore, it is troubling that Items 2, 3, and 41 have negative correlations after being reverse scored. Either these items do not function as intended or they pinpoint

Table 2

CSQ First Pilot Item Fit Statistics

Item Number	Item-Total Correlation	Outfit		Item Difficulty
		Mean-square	Standardized	
1	.35	1.09	1.00	0.19
2[a]	-.06	1.52	7.30	-0.62
3[a]	-.15	1.95	9.90	-0.89
4	.26	0.81	-3.00	-0.34
5	.45	1.05	0.60	0.67
6	.43	0.91	-1.10	0.12
7	.47	0.93	-0.80	0.61
8	.48	0.96	-0.40	0.81
9	.31	0.98	-0.20	-0.26
10[a]	.20	1.05	0.70	-0.26
11	.44	0.85	-1.70	0.64
12	.39	0.77	-3.00	0.17
13	.34	0.95	-0.60	-0.12
14	.34	1.11	1.30	0.26
15	.48	0.76	-3.20	0.18
16	.41	0.72	-3.60	0.28
17	.47	0.77	-2.80	0.42
18	.37	1.00	0.00	0.17
19	.30	0.98	-0.30	-0.32
20	.43	0.77	-3.10	0.04
21	.30	1.00	0.00	0.13
22[a]	.08	1.93	9.50	0.01
23	.42	0.77	-3.10	0.15
24	.37	0.97	-0.30	0.01
25	.42	0.73	-3.50	0.24
26	.39	0.77	-3.10	0.10
27[a]	.09	1.07	1.10	-0.44
28	.36	0.87	-1.80	0.05
29	.33	0.83	-2.20	0.11
30	.41	0.81	-2.60	0.04
31	.46	0.78	-2.70	0.37
32	.50	0.69	-4.40	0.05
33	.33	0.90	-1.30	0.20
34	.52	0.81	-2.30	0.32
35	.35	0.96	-0.50	-0.05

(Table continues)

Table 2 (continued)

CSQ First Pilot Item Fit Statistics

Item Number	Item-Total Correlation	Outfit Mean-square	Outfit Standardized	Item Difficulty
36[a]	.12	1.73	8.70	-0.25
37[a]	.05	1.18	2.70	-0.50
38	.44	0.67	-4.40	0.32
39[a]	.05	1.29	4.30	-0.52
40	.39	0.68	-4.30	0.26
41[a]	-.27	1.63	7.30	-1.25
42[a]	.12	1.30	4.10	-0.36
43[a]	.05	1.34	4.60	-0.37
44	.49	0.66	-4.80	0.12
45[a]	.10	1.40	5.60	-0.50

[a] These items were eventually deleted from the questionnaire.

a misconception about the NOS held by many students. Items 2, 3, and 41 are presented here for the reader's convenience:

> 2. Scientific research methods can be used to make valid judgments about moral issues.
>
> 3. There are no limits to the types of questions that can be asked and answered by science.
>
> 41. The same step-by-step procedure commonly known as the 'Scientific Method' is used by scientists in all fields of research.

Items 2 and 3 refer to the types of questions that can be answered by science. High school students may believe that scientific inquiry is capable of answering questions of a moral or spiritual nature. Item 41 highlights a misconception commonly held by secondary students who are not knowledgeable about the nature of scientific inquiry.

Rasch model statistics. After considering item-total correlations, the outfit mean-square statistic for each item is examined. The outfit mean-square is the sum of squared residuals of each item to the Rasch model prediction. This is essentially a chi-square statistic divided by its degrees of freedom (Linacre & Wright, 1994). The outfit statistic is particularly sensitive to outlying responses which do not fit the expected response pattern. Such responses generally yield little useful information (Linacre & Wright, 1994). The expected value of the outfit statistic is 1.0. Values less than 0.5 or greater than 1.5 indicate items for which the expected response as predicted by the model is significantly different than the actual response given. Values greater than 2.0 are particularly damaging to the estimates of model fit statistics (Wright & Linacre, 1994).

Deciding which items to exclude and which to retain based on fit statistics is a complicated process. One examines the outfit mean-square value and item-total correlation for the item, considers what was intended by the item, and then makes an informed judgment whether to retain, exclude, or revise the item. Once an item or group of items has been excluded, the analysis is repeated with the remaining items. One danger of repeating this process too much is that all of the items will eventually be deleted. As items are excluded, the new model and fit statistics are based on the new item set; therefore, it is important to retain in memory the original fit statistics from the initial analysis and to use these as a guide in further decisions.

Examining the outfit mean-square values in Table 2 shows that Items 3, 22, and 36 have outfit mean-square values greater than 1.5 and nearly as great as the 2.0 cut-off value. The remaining items have outfit mean-square values well below the cutoff value of 2.0 and are all within the range of 0.5 to 1.5. These items should be considered for retention.

Item difficulty. Finally item difficulty must be considered. If especially difficult or especially easy items are deleted, this may skew the questionnaire results. In order to determine what students really know, a spread of item difficulties from hard to easy is desirable. Figure 1, obtained from Winsteps, shows the person ability estimates ordered from greatest ability at the top to least ability at the bottom on the left of the vertical line and the item difficulty measures ordered from most difficult to endorse at the top to easiest to endorse at the bottom on the right of the vertical line. Each period (.) on the left represents one student at that ability level and each cross-hatch (#) represents six students at that ability level. *M* locates the median value of the distribution while *S* and *T* locate one and two standard deviations respectively.

```
1                    +
                     |
                  .  |T Q8
                     |  Q11   Q5    Q7
                     |
                  .  |S Q17   Q31   Q34   Q38
                  .  |  Q1    Q14   Q16   Q25   Q33   Q40
                 .# T|  Q12   Q15   Q18   Q21   Q23   Q26   Q29   Q44
      Q6
0                ##  +M Q20   Q22n  Q24   Q28   Q30   Q32   Q35
               .### S|  Q13
         .#########  |  Q10   Q36n  Q9
        ###########  |S Q19   Q4    Q42n  Q43
         ########## M|  Q27   Q37   Q39n  Q45
             .#####  |  Q2n
            .###### S|T
               .###  |  Q3n
-1               .#  +
                  # T|
                  .  |  Q41n
                  .  |
                  .  |
                  .  |
                  .  |
                     |
-2                   +
```

Figure 1. CSQ First Pilot Item Difficulty and Person Ability Map.

The distribution of questions is quite symmetric showing a slight skew towards the easier to endorse items. The person ability estimates are also symmetric but are shifted down relative to the item difficulties. This means that it was relatively difficult for these students to express agreement or disagreement to these items. Items which have been previously identified for deletion are shown in Figure 1 in a boldface, slightly larger font. Notice that these items are predominantly at the easy end of the item difficulty scale. Simply deleting them from the questionnaire would skew the items towards more difficult items.

Internal Consistency Reliability

Internal consistency reliability is most commonly estimated using Cronbach's coefficient alpha. Another way of calculating internal consistency reliability estimates is through Rasch analysis (Tedman & Keeves, 2001). Estimation of internal consistency reliability coefficients assumes that individual subscales are unidimensional (Netemeyer et al., 2003).Cronbach's alpha and the reliability estimates given by Rasch analysis are analogous in interpretation, although the Rasch estimate is generally slightly smaller.

When responses were first submitted to Rasch analysis utilizing the Winsteps program, the Cronbach's alpha for all 45 items (N=353) was α =.79. At first blush, this seems low for a scale including 45 items. Analysis of item correlation and fit statistics revealed that several items did not function well. When items which were worded with a negative orientation were excluded from the analysis, Cronbach's alpha increased to α =.87. This is a marked increase in reliability. Further analysis suggested excluding several other items from analysis. Removing these items yields α =.88. These results are summarized in Table 3. Deleting more items does not improve estimates of internal consistency reliability. However, deleting items may have serious affects on the validity of the questionnaire relative to ILO 6. Therefore, items were revised whenever possible.

Table 3

Internal Consistency Reliability Estimates for Three Scale Configurations

Item Configuration	Specific Items Excluded	Alpha
All Items	None	.79
Negative Items Excluded	2, 3, 22, 36, 39, 41, 42	.87
Other Items Excluded	10, 27, 37, 43, 45	.88

In conclusion, the first pilot of the CSQ revealed several items, particularly the negatively oriented items, which were deleted or revised. All items were examined with an eye critical to wording, vocabulary, and meaning. Items which were vague or misleading were reworded regardless of their statistics as presented in Table 2. With consideration given to the item map in Figure 1, the items were primarily revised. Items which could not be reasonably revised were deleted and rewritten. This left 33 items in the questionnaire.

Response Category Functioning

The response categories for the first version of the CSQ were 0 (*strongly agree*), 1 (*agree*), 2 (*slightly agree*), 3 (*slightly disagree*), 4 (*disagree*), and 5 (*strongly disagree*). Figure 2 shows the response category function curves for the retained items on the questionnaire. The intersections of the lines indicate the points where the response options become equally endorsable. Ideally curves 1 through 4 should be of uniform height.

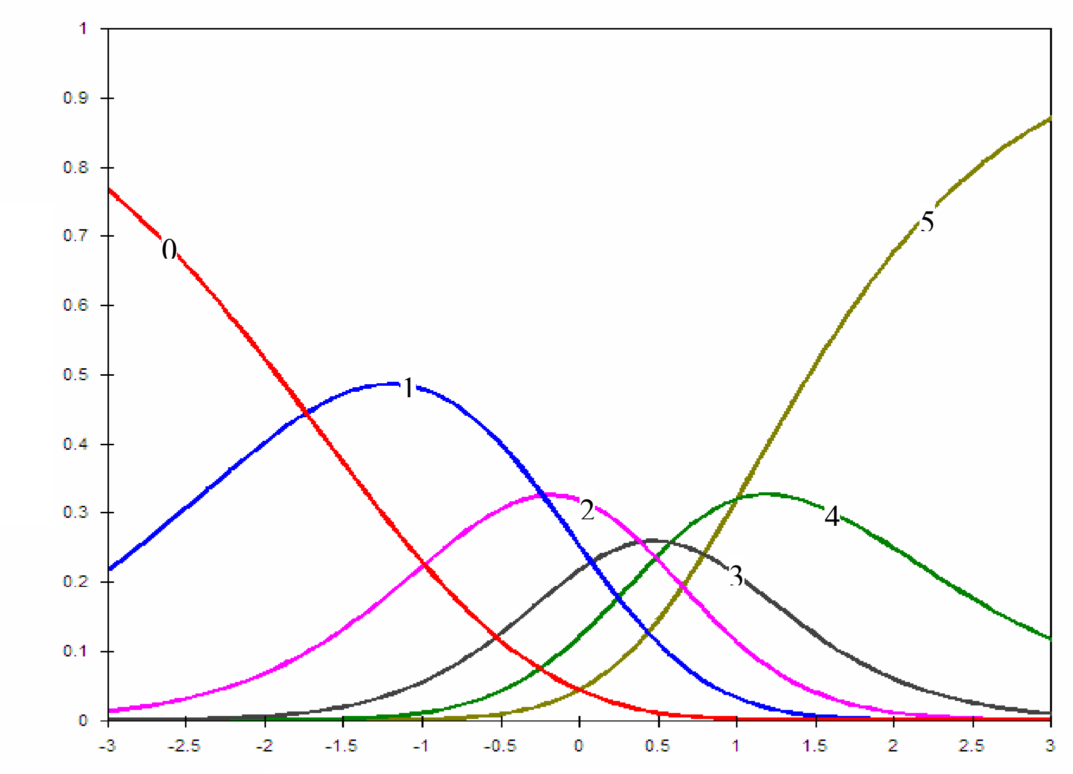

Figure 2. Category Probabilities for the CSQ.

Response 3 (*slightly disagree*) intersects both responses 2 (*slightly agree)* and 4 (*disagree*) at nearly the same point on the trait scale. This indicates that response 3 did not function well on this questionnaire. Respondents were less likely to choose this response than to choose either response 2 or response 4. In other words, students taking the questionnaire were unable to distinguish between slightly disagree as compared to either slightly agree or disagree.

The category response statistics are shown in Table 4. These statistics indicate that option 5 (s*trongly disagree*) was not a popular choice. This could be due to the positive orientation of the majority of the items, but it also indicates that this option did not function as intended. Furthermore, the structure calibration column indicates that not all response categories are needed. The structure calibration values increase monotonically as expected. The calibration structure values should also demonstrate a minimum 0.9 logit spread between categories when 6 categories are used. The spread between categories 0-1, 1-2, and 4-5 are satisfactory; however, the spread between categories 2-3, and 3-4 do not meet this criterion.

Table 4

First Pilot Category Structure Diagnostic Statistics

	Observed			Expected	Structure
Response Category	Count	%	Average	Average	Calibration
0 Strongly Agree	2337	20	-1.36	-1.36	-3.00
1 Agree	4543	39	-0.90	-0.88	-1.21
2 Slightly Agree	2788	24	-0.56	-0.60	-0.18
3 Slightly Disagree	1172	10	-0.33	-0.37	0.48
4 Disagree	509	4	-0.19	-0.14	1.21
5 Strongly Disagree	181	2	-0.06	0.15	2.49

The category structure calibration values, response frequency counts, and indications from the category probabilities shown in Figure 2 give evidence that two or more of the response categories should be collapsed

Figure 3 shows the probability curves after combining response categories into only four options. The newly formed category probabilities appear somewhat lopsided, but the two center categories are mostly the same height and the peaks are generally evenly spaced. This provides evidence that the four revised categories would likely

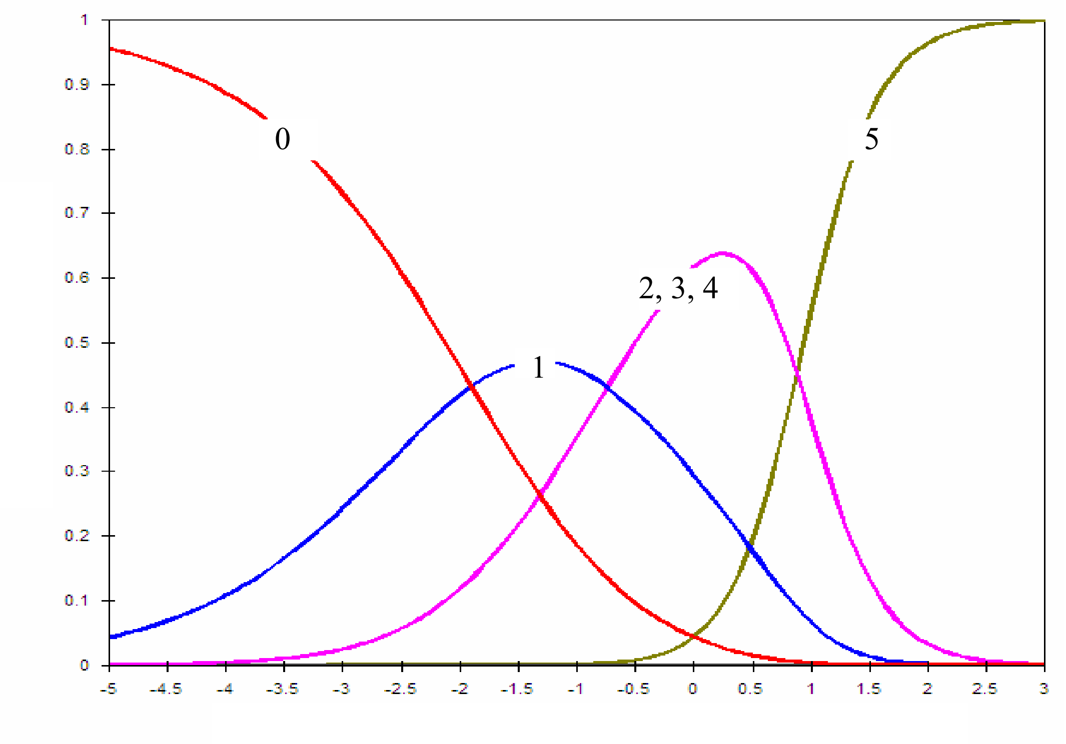

Figure 3. Category Probabilities After Combining Response Categories.

function better than the original six. Figure 3 also indicates that students responding to the items on the first pilot were unable to adequately distinguish the differences in meaning between the six responses which were originally supplied.

Refinement and Re-administration

Based on the analysis of data from the pilot test, items were refined and deleted from the initial version of the CSQ. A new scale was developed from the refined items and administered to a group of secondary students similar to those who participated in the pilot test. Teachers from a different high school located near the original were asked to administer the revised version of the CSQ to their students. The researcher believed that the two groups of students from geographically close schools would represent the same population. This method provides enough parallelism between groups to establish the reliability and validity of the assessment. The refined questionnaire was administered in the spring of 2007.

Administration of the Revised CSQ

Items which were retained from the first pilot administration were revised and augmented with new items. The number of response options for the revised CSQ was reduced from the six options mentioned previously to four new options. The new response options were 0 (*definitely true*), 1 (*probably true*), 2 (*probably false*), and 3 (*definitely false*). Items in the revised CSQ were reviewed by a panel of experts and by members of the target population prior to administration of the questionnaire. The new version of the questionnaire including 38 items was administered to 545 secondary students in the spring of 2007. As before, responses were analyzed with a combination of Rasch methods and more traditional techniques. Fit statistics for the second administration of the CSQ are presented in Table 5.

Item Retention, Revision, or Deletion

Item-total correlation. Examining the item-total correlation values in the second column of Table 5 shows the following items with correlation values below .30: Items 1, 2, 3, 4, 5, 6, 8, 9, 10, 13, 15, 16, 18, 19, 20, 21, 22, 24, 27, 29, 30, 31, 32, 35, 36, 37, and 38. Several of these items have correlation values between .20 and .30 and so could be considered moderately acceptable. This would allow the re-inclusion of Items 2, 3, 13, 15, 16, 22, 24, 30, 32, 35, and 37; leaving 22 items in the questionnaire.

Note that all negatively worded items are included on the initial list of potential items for deletion. Negative items may function as a separate scale (Nielson, 2002). To check this possibility, the positive item total was correlated with the negative item total

Table 5

Revised CSQ Item Fit Statistics

Item[a]	Item-Total Correlation	Outfit Mean-square	Outfit Standardized	Item Difficulty
1	-.02	1.27	4.70	-0.17
2	.22	0.99	-0.10	0.17
3	.21	1.10	1.30	1.09
4N	.05	1.11	2.30	-0.64
5	.05	0.95	-0.90	-0.31
6	.14	0.90	-2.10	-0.40
7	.34	1.14	1.90	0.84
8	-.02	1.25	5.00	-0.84
9	.10	0.90	-1.90	-0.30
10	.16	0.92	-1.30	0.08
11	.36	1.04	0.50	1.03
12	.33	0.82	-3.10	0.33
13	.27	1.14	2.10	0.53
14	.39	0.87	-2.10	0.53
15	.24	0.83	-3.20	-0.08
16N	.25	1.17	3.30	-0.28
17	.30	0.88	-1.90	0.54
18	.06	0.83	-3.60	-0.54
19	.18	1.01	0.20	-0.37
20	-.05	1.24	4.70	-0.96
21N	-.17	0.99	-0.20	-1.20
22	.29	0.91	-1.50	0.38
23	.39	1.13	1.80	0.61
24N	.26	1.36	5.00	0.62
25	.39	0.77	-3.90	0.33
26	.33	0.90	-1.60	0.43
27N	.17	0.88	-2.50	-0.24
28	.34	0.84	-2.80	0.26
29	-.05	1.09	1.80	-0.90
30	.29	0.99	-0.10	0.01
31	-.06	1.22	4.20	-0.67
32	.28	1.07	1.20	-0.10
33	.31	1.01	0.20	-0.15
34	.34	0.87	-2.20	0.37
35	.26	0.87	-2.00	0.40

(Table continues)

Table 5 (continued)

CSQ Second Pilot Item Fit Statistics

	Item-Total	Outfit		Item
Item[a]	Correlation	Mean-square	Standardized	Difficulty
36N	.12	0.90	-2.20	-0.47
37	.26	1.15	2.30	0.23
38	.04	0.93	-1.20	-0.17

[a] Items identified with 'N' were negatively worded and reverse scored.

yielding a correlation coefficient of .091. A Pearson correlation value this low indicates that indeed the negatively worded items do not function in the same way as the positively worded items and should be excluded from further analysis. Removing only the negatively oriented items would leave 32 items in the questionnaire.

Rasch model statistics. As can be observed in Table 5, all of the items in the second version of the CSQ have outfit mean-square values within the acceptable range of 0.5 to 1.5. Therefore the items in the revised CSQ may not function as poorly as the item-total correlation values might suggest. It is likely that the poor item-total correlation values may be due to multidimensionality of the NOS.

Item difficulty. Figure 4 charts the item difficulty and person ability measurements in a manner similar to Figure 1. Once again a period (.) represents one student and a crosshatch (#) represents six students. As can be seen in Figure 4, the items in the revised CSQ were relatively difficult for this sample of students. The distribution of items is mostly symmetrical, exhibiting some bimodal behavior. The items are spread over the range of difficulty, with most grouped near the center of the item difficulty scale, but exhibiting a slight skewness towards the easier items.

```
 2                +
                  |
                  |
                  |
                  |
                  |
                  |
                  |
                  |
                  |T Q3
 1                +  Q11
                  |
                  |  Q7
                  |
                  |S Q23    Q24N
                  |  Q13    Q14    Q17
                  |  Q22    Q26    Q34    Q35
               .  |  Q12    Q25    Q28
               .  |  Q2     Q37
               . T|  Q10
 0           .##  +M Q30
             .##  |  Q15    Q32
          .#####  |  Q1     Q27N   Q33    Q38
        .####### S|  Q16N   Q5     Q9
        .#######  |  Q19    Q6
     .##########  |  Q18    Q36N
   .############ M|S Q4N
         .######  |  Q31
       .########  |  Q8
     .##########  |  Q29
-1         .#### S+  Q20
            .###  |T
              .#  |  Q21N
             .##  |
              .# T|
               .  |
               .  |
               .  |
               .  |
                  |
-2                +
```

Figure 4. Revised CSQ Item Difficulty and Person Ability Map.

Since the negative items do not function in the same manner as the positive items, they are first to be considered for deletion. Figure 4 shows that the negative items (labeled with an *N* following the item number) are fairly evenly disbursed among the other items, therefore deleting them would not likely affect the difficulty of the CSQ.

Validity considerations. Before making any recommendations concerning item deletion, a look at the alignment of the items with the indicators of ILO 6 is warranted. Table 6 shows the alignment of CSQ items to the indicators of ILO 6. Items which have been identified for deletion or revision are marked with a superscript b (^{b}). Table 6 shows that if the indicated items are deleted, then several indicators would be underrepresented in the questionnaire. Since this is undesirable, those items should be rewritten if possible.

Internal Consistency Reliability

With 38 items the initial value of Cronbach's alpha was quite low at α =.66. As mentioned previously, the negatively oriented items do not appear to function well with the positively oriented items. Removing the negative items from the analysis increased the internal consistency reliability to α =.68, still much lower than the internal consistency reliability estimates of the first pilot. The Spearman-Brown prophecy formula predicts an internal consistency reliability estimate of .84 for the revised CSQ after deleting negative items. However, if the items of the CSQ reflect an underlying multidimensionality of the NOS, then the internal consistency reliability of all the items together should be expected to be low. Other factors which may help explain the low internal consistency reliability estimate of the revised CSQ are offered in Chapter 5. While many items have low item-total correlations, they all have acceptable outfit mean-square values. Deleting a large number of items with the intent of increasing internal

Table 6

Alignment of Questionnaire Items to the Indicators of ILO 6

	Indicator	Items
A.	Science is a way of knowing that is used by many people, not just scientists.	2, 28
B.	Understand that science investigations use a variety of methods and do not always use the same set of procedures; understand that there is not just one “scientific method.”	4[b], 9[b]
C.	Science findings are based on evidence.	1[b], 5[b], 16[b], 19[b], 25, 26
D.	Understand that science conclusions are tentative and therefore never final. Understandings based upon these conclusions are subject to revision in light of new evidence.	3, 8[b], 12, 24[b], 31[b], 37
E.	Understand that scientific conclusions are based on the assumption that natural laws operate today as they did in the past and that they will continue to do so in the future.	32, 33
F.	Understand the use of the term *theory* in science, and that the scientific community validates each theory before it is accepted. If new evidence is discovered that the theory does not accommodate, the theory is generally modified in light of this new evidence.	10[b], 13, 20[b], 29[b]
G.	Understand that various disciplines of science are interrelated and share common rules of evidence to explain phenomena in the natural world.	35, 38[b]
H.	Understand that scientific inquiry is characterized by a common set of values that include logical thinking, precision, open-mindedness, objectivity, skepticism, replicability of results, and honest and ethical reporting of findings. These values function as criteria in distinguishing between science and non-science.	15, 23
I.	Understand that science and technology may raise ethical issues for which science, by itself, does not provide solutions.	21[b], 30

(Table continues)

Table 6 (continued)

Alignment of Questionnaire Items to the Indicators of ILO 6

	Indicator	Items
J.	Science is a public endeavor requiring peer review and open, honest reporting of findings.[a]	7, 11, 14, 17, 22, 34
K.	Science is subject to the experience, history, perceptions, emotions, opinions and views of scientists. Scientific inquiry is a human endeavor.[a]	6[b], 18[b], 27[b], 36[b]

[a] These two statements are not in ILO 6 but are suggestions to fill a perceived inadequacy for which items were written.

[b] Denotes items under consideration for deletion or refinement.

consistency reliability estimates may have the effect of obscuring meaningful information about the NOS construct.

Scale Structure

The nature of science is complex and has frequently defied description (Lederman, Wade, & Bell, 1998; McComas et al., 1998; McComas & Olson, 1998). An exploratory factor analysis using the principal component method and oblimin obliquerotation was used to reveal the structure underlying the NOS as described in the questionnaire. Negatively worded items were not included in the principal components factor analysis. The results of this procedure are presented in Table 7. There are 11 components with eigenvalues greater than one. Together these components account for 53.2% of the observed variance in the responses to the questionnaire. A scree plot of these results is shown in Figure 5.

The scree plot and eigenvalues indicate one strong component accounting for 12.8% of the variance in the responses with the second and third components accounting

Table 7

Eigenvalues for the Principal Components

Component	Eigenvalue	% Variance Explained	Cumulative % Variance
1	4.1	12.8	12.8
2	1.9	6.0	18.8
3	1.5	4.7	23.5
4	1.4	4.4	28.0
5	1.3	4.1	32.1
6	1.3	4.0	36.1
7	1.2	3.7	39.8
8	1.2	3.6	43.4
9	1.1	3.4	46.9
10	1.0	3.2	50.1
11	1.0	3.1	53.2

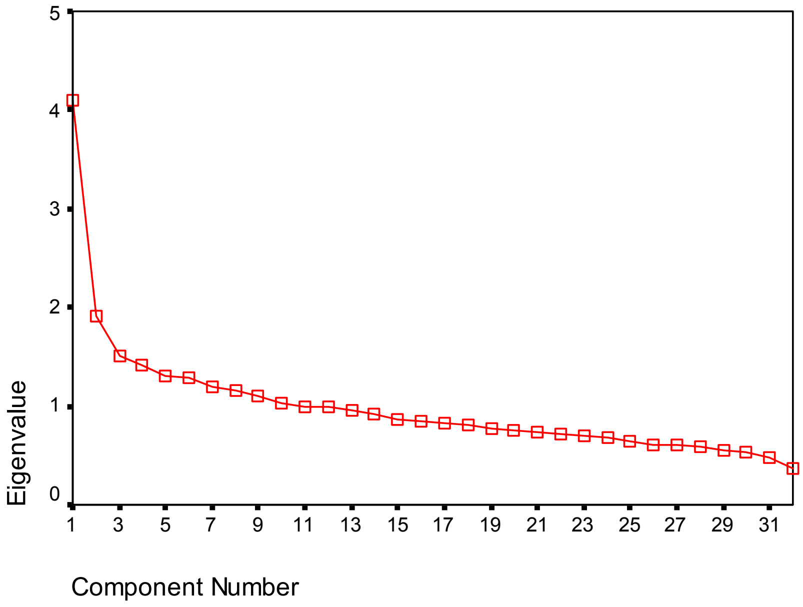

Figure 5. Scree Plot of Principal Components.

for 6.0% and 4.7% of the variance respectively. The remaining components each account for 3% to 4.5% of the total variance.

The factor pattern coefficients matrix for the 11 rotated components is shown in Table 8. Items which load on more than one factor will be considered first, then the components will be examined. Items are presented in the same order in which they appear in Table 8.

Dual-loading items should be assigned to one component if possible. To decide which component to place an item with, the items in each component were examined together to search for common meaning and intent. Item 7 loads on Components 1 and 8, but seems to share meaning with the items of Component 1. Item 8 loads on Components 3 and 8 but will be included with Component 3. Item 20 loads onto Components 3 and 4 but aligns best when included with Component 4. Item 38 loads onto Components 1 and 6 but will be included with Component 1 for further analysis. Item 26 loads on Components 6 and 7 but shares common meaning with the items of Component 6. Item 6 loads onto Components 3 and 7 but will be included with Component 7. Item 9 loads on both Component 6 and Component 9 but loads onto Component 9 much more strongly. Item 25 loads onto Components 6 and 9 and will be included with Component 6. Item 13 loads onto Components 10 and 11 and it will be included with Component 10 for further analysis. Additionally, Items 28 and 30 do not load strongly onto any of the components.

The components shown in Table 8 after deciding where to place dual-loading items remain as follows. Component 1 contains Items 7, 11, 17, 22, 23, and 38. Component 2 includes Items 32 and 33. Component 3 consists of Items 8, and 31. Component 4 includes Items 1, 19, and 20. Component 5 consists of Items 3, 10, and 14.

Table 8

Revised CSQ Factor Pattern Matrix from Oblimin Rotation[a]

	Component										
Item	1	2	3	4	5	6	7	8	9	10	11
22	.53										
23	.51										
17	.50										
7	.45							.33			
11	.41										
32		.86									
33		.83									
31			.67								
8			.61					-.40			
20			.54	.31							
30											
1				.81							
19				.71							
10					.65						
3					.56						
14					.32						
29						.70					
38	.46					.53					
26						-.47	.35				
15							.70				
35							.51				
6			.34				.37				
28											
2								.70			
12								.38			
9						.34			.80		
37									.47		
25						-.37			.39		
5										.87	
13										.45	-.42
18											.73
34											-.43

[a] Loading values less than 0.3 have been suppressed to allow for easier visual analysis of the components.

Component 6 includes Items 25, 26, and 29; note that Items 25 and 26 load negatively onto this component. Component 7 contains Items 6, 15, and 35. Component 8 includes Items 2 and 12. Component 9 has Items 9 and 37. Component 10 includes Items 5 and 13. Component 11 consists of Items 18 and 34. Interpreting the components may be possible if the items which load on those components have meaningful similarities.

Comparing Table 8 to Table 6 reveals the complex structure of the nature of science. The items of the CSQ do not load onto the rotated components in the same groupings shown in Table 6. The indicators of ILO 6 described in Table 6, which were used as an initial domain map for item writing, likely overlap conceptually with each other in describing the nature of science.

The reliability estimates for the items grouped according to each component are shown in Table 9. Given the relatively low number of items in any component, the reliabilities are expected to be low, and the values shown in Table 9 indicate that this is true. Note that the reliability estimates for Components 6 and 11 each have negative values. This undesirable result is a consequence of having two items on each of these scales. In each of Components 6 and 11 the two items are negatively correlated with each

Table 9

Revised CSQ Factor Analysis Component Reliabilities

	Component										
	1	2	3	4	5	6	7	8	9	10	11
Reliability	.58	.73	.39	.37	.33	-.16	.27	.30	.23	.30	-.08

other. The reliabilities of the 11 components could likely be improved by writing new items that are meaningful in the context of the items already included in each factor.

Response Category Functioning

Table 10 summarizes the category functioning statistics obtained from the Winsteps program for the items grouped as indicated in Table 8. The category functioning statistics for all 11 components exhibit close agreement between the observed and expected averages for each response category, except response 3 (*definitely false*). In this case many of the observed averages are smaller than the expected averages. The category measure values increase monotonically with no reversals. Additionally, there are 1 to 1.5 logits between category measures. However, the step calibration values reveal that some response categories may not function well for all subscales. In particular response 2 (*probably false*) does not appear to function well as seen in Factors 1, 2, 4, 5, 6, 7, and 9. In all of these factors, the step threshold values between response 2 (*probably false*) and response 3 (*definitely false*) are relatively small. This indicates that students were less able to distinguish between these two responses than between the others and that perhaps the response options ought to be combined in some way.

Table 10

Category Structure Statistics with Items Grouped as Shown in Table 8

	Response	Observed		Expected	Category	Step
Factor	Category	Percent	Average	Average	Measure	Threshold
1	0	51	-1.35	-1.31	-2.09	None
	1	33	-1.00	-1.06	-0.55	-0.75
	2	11	-0.78	-0.82	0.59	.14
	3	3	-0.64	-0.59	2.03	.61
2	0	32	-.072	-0.66	-2.07	None
	1	38	-0.46	-0.49	-0.53	-0.74
	2	20	-0.29	-0.35	0.58	0.22
	3	9	-0.22	-0.23	1.98	0.52
3	0	17	-0.05	-0.21	-2.08	None
	1	30	0.03	-0.02	-0.59	-0.69
	2	34	.11	.13	0.57	-0.06
	3	20	.09	.27	2.10	0.75
4	0	21	-0.38	-0.45	-2.20	None
	1	38	-0.20	-0.21	-0.54	-0.91
	2	25	0.04	0.02	0.63	0.35
	3	16	0.11	0.24	2.05	0.57
5	0	50	-1.43	-1.39	-2.22	None
	1	37	-1.03	-1.08	-0.54	-0.94
	2	10	-0.79	-0.81	0.64	0.36
	3	3	-0.63	-0.60	2.06	0.58
6	0	32	-1.11	-1.09	-2.38	None
	1	40	-0.71	-0.69	-0.61	-1.12
	2	19	-0.03	-0.18	0.70	0.30
	3	9	0.09	0.26	2.22	0.82
7	0	29	-0.97	-0.93	-2.50	None
	1	47	-0.62	-0.64	-0.60	-1.28
	2	18	-0.34	-0.38	0.76	0.49
	3	6	-0.26	-0.17	2.24	0.79

(Table continues)

Table 10 (continued)

Category Structure Statistics with Items Grouped as Shown in Table 8

	Response	Observed		Expected	Category	Step
Factor	Category	Percent	Average	Average	Measure	Threshold
8	0	40	-1.14	-1.11	-2.36	None
	1	42	-0.93	-0.94	-0.63	-1.08
	2	15	-0.75	-0.79	0.69	0.18
	3	3	-0.63	-0.66	2.27	0.90
9	0	31	-0.86	-0.82	-2.27	None
	1	42	-0.54	-0.59	-0.55	-1.00
	2	18	-0.36	-0.38	0.66	0.39
	3	7	-0.40	-0.21	2.09	0.61
10	0	37	-1.12	-1.08	-2.27	None
	1	39	-0.69	-0.77	-0.62	-0.96
	2	19	-0.46	-0.49	0.65	0.11
	3	6	-0.55	-0.26	2.22	0.85
11	0	28	-1.15	-1.10	-2.60	None
	1	44	-0.66	-0.70	-0.72	-1.36
	2	22	-0.26	-0.32	0.79	0.19
	3	6	-0.27	-0.03	2.49	1.17

Chapter 5: Conclusions and Recommendations

Creating a questionnaire to assess students' beliefs about the nature of science is a challenging task. This task is made more challenging by the lack of consensus among scholars regarding the characteristics of scientific inquiry and knowledge. However, within the limited scope of the Utah Core Curriculum in Science, the questionnaire developed for this project may allow secondary teachers to assess their students' beliefs about the nature of scientific inquiry and knowledge.

Conclusions

Validity and Reliability of the CSQ

The conclusions presented in this chapter are based upon responses given by secondary students during the administration of the revised CSQ. Several factors arising from personal experience over many years teaching secondary-age students cause me to exercise some caution in forming these conclusions. First, this age group of students is already disinclined to respond seriously to questionnaires. Second, as required, students and parents signed a release form. Since the researcher named on the release form was a person to whom the students were not accountable in any way, this may have caused them to treat the items and their responses lightly. Wolf, Smith and Birnbaum (1995) assert that high school students will behave on a test in accordance with their perception of personal benefit to be gained from performance. Third, the teachers administering the questionnaire were not accountable to me, the researcher, nor were some even well acquainted with me. I believe that adults as well as high school students tend to behave in accordance with the perceived benefit to be gained from a particular activity. Personal communication with a teacher at the high school where the revised questionnaire was administered indicated that other teachers may not have encouraged students to do their best, but rather treated the questionnaire administration as an intrusion. Fourth, student responses to the questionnaire may actually reveal misconceptions and a lack of direct instruction in the nature of science. Any one of these factors could cast some doubt on the results; taken together I believe that they at least partially account for the lack of inter-item consistency indicated by the low reliability estimates.

Validity. The nature of science appears to be multi-dimensional. Evidence for this claim is shown by the results of the factor analysis summarized in Table 8. Internal consistency reliability is high when the inter-item correlations are high. High inter-item correlation requires that the items of a scale behave as though they were a single measure. Dawis (2000) states, "Obviously, the internal consistency idea of reliability cannot apply to heterogeneous scales" (p. 87). As mentioned previously, there is a trade-off between reliability and validity. In order to enhance the validity of the CSQ, students were necessarily asked to respond on a multidimensional scale. Low reliability estimates should have been anticipated by the researcher.

The CSQ was designed with external validity in mind. Items were written specifically to address the indicators of ILO 6. Students from the target population were asked to review the items for vocabulary and interpretation. Their input was given great weight when editing the items prior to administering the CSQ. Furthermore, the items were rated by individuals knowledgeable about science education, ILO 6, and the nature of science. They gave meaningful critiques of the suitability of the items for secondary students as well as the alignment of the items to the indicators of ILO 6. However, because the internal consistency reliabilities of the CSQ are low, the ability of the CSQ to measure the target knowledge of the NOS expressed in ILO 6 is tenuous.

The validity of the CSQ may be further obscured by the possibility that the students participating in the pilots had not been sufficiently instructed in the NOS. If the CSQ has external validity to ILO 6, it would be difficult to demonstrate that based on the responses of students who do not have correct understanding of the NOS. Inconsistent response behavior by students would likely reduce the inter-item correlations within components which would in turn reduce the reliability estimates of the CSQ.

Reliability. Dawis (2000) asserts,

> Reliability should not be thought of only as a scale characteristic; it is also affected by the respondent sample. That is, reliability is correctly seen as an attribute – not of the scale – but of the scale scores, and scale scores are the product of scale-by-respondent interaction. (p. 87)

In other words, reliability is dependent upon the way that students perceive the items, the instruction students have received about the topics covered by the questionnaire, the way in which the questionnaire was presented to them, the attitude of the person administering the questionnaire, and the attitudes of students towards questionnaires in general, among other factors. Most of these conditions were present when students responded to the CSQ and I believe that the reliability obtained from this administration was affected by these conditions.

The lack of direct instruction in the NOS is of particular concern. If students have not been instructed in the nature of science, their answers to the items would be random guesses at best. As a result students may have been led to select a default response (such as *definitely true*), defined as a response set, and to mark that response on a large number of items without consideration to the content of the item. Schulz and Sun (2001) define this behavior as an undifferentiating response set. They found that such response sets decrease overall reliability. No clear evidence of response set behavior could be found from the data; however, increased response randomness would likely increase random error in the data which would decrease the reliability of the questionnaire.

Furthermore, Alwin and Krosnick (1991) report that lack of reliability is widely attributed to a lack of education of the respondents. I believe that the students who responded to the revised CSQ may not have been instructed specifically in the nature of science. I make this statement for several reasons. First, my work on this project has revealed my own misconceptions. I was never specifically instructed in NOS throughout my teacher-education coursework nor in any of my studies of the sciences until this project. Second, conversations with other science teachers in my department and district reveal that many of them hold several of the same misconceptions that I did. As my knowledge has grown, I have become attuned to the misunderstandings held by others. Third, current research reveals a lack of specific instruction in NOS at many institutions responsible for educating science teachers (see Abd-El-Khalick, 2000, 2005; Clough, 1998; Lederman, 1992; Martin-Diaz, 2006; McComas, Clough, & Almazroa, 1998; Smith & Scharmann, 1999; and Tedman & Keeves, 2001 for examples). Teachers who have not been instructed in the NOS are unlikely able to teach it effectively to their own students.

In spite of these potential shortfalls, the information obtained from the revised questionnaire provides useful information about the items and the questionnaire that can be used for further refinement. Although it is beyond the scope of this project, I am not finished with this questionnaire. This study takes a meaningful step towards allowing Utah science teachers to assess what their students believe about the nature of the scientific enterprise. An instrument aligned to the Utah Core Curriculum in Science, such as the one developed in this project, did not exist prior to this project. This questionnaire can open a dialogue among science teachers concerning how the nature of science can be more effectively taught.

Item Retention, Revision, or Deletion

Analysis reveals that negatively worded items do not function in the same manner as positive items (Nielson, 2002); therefore, the negatively worded items in the CSQ should be reworded as positive statements. To avoid undercoverage of the indicators of ILO 6, items will be reviewed for revision prior to being deleted. Also, in an effort to increase the reliability of the CSQ, new items should be written to augment the number of items covering each component shown in Table 8.

Dimensional Structure

Because there are eleven indicators in ILO 6, I initially expected up to eleven components which aligned with the indicators of ILO 6. While 11 components were obtained, they do not simply mimic the indicators of ILO 6. Apparently each of the indicators specified in ILO 6 does not describe a unique aspect of the NOS; rather, the indicators overlap conceptually.

The following paragraphs describe the components obtained from the factor analysis. The components identified in Table 8 might describe meaningful aspects of the nature of science. However, caution must be exercised in placing too much confidence in the interpretations presented here. The reliability estimates from the set of responses to the CSQ on which principal component factor analysis was performed are low (see Table 9). In order to preserve some degree of external validity, all but the negatively worded items were included in the factor analysis. Some attempt to draw meaning from the components may provide enlightenment and seems worthwhile as long as it is done cautiously.

Component 1. The first component consists of Items 7, 11, 17, 22, 23, and 38. The items of this component describe the scrutiny to which the results of scientific inquiry are subjected. Scientists have the responsibility to report their own work in sufficient detail that others may judge the adequacy of the methods used and the conclusions drawn. These items also express the manner in which questions raised by previous work form the basis of new inquiries. They also state that collaboration and review by scientists from different disciplines enhances the quality of their work and their productivity.

Component 2. Items 32 and 33 load onto the second component. This is not surprising since these two items were both written specifically to address Indicator E of ILO 6 about the assumption that natural laws operate today as they did in the past and that they will continue to operate in the same manner in the future. Interestingly, these two items have grouped exclusively on the same component no matter which other items were included or excluded from the factor analysis.

Component 3. Items 8 and 31 load on this component. These two items express the inability of scientific inquiry to ever establish absolute truth with finality.

Component 4. Component 4 of the factor analysis consists of Items 1, 19, and 20. Together these items express the empirical nature of scientific inquiry and knowledge. Results are often verified by scientists working independently.

Component 5. The fifth component identified in the factor analysis consists of Items 3, 10, and 14. These items are more difficult to interpret; however, they seem to relate to the tentative and cyclical nature of the development of scientific knowledge.

Component 6. This component includes Items 25, 26, and 29. Items 25 and 26 state that the quality of the explanations scientists make to describe recurring patterns depends on the quality of the data to begin with. However, Item 29 expresses the concept of parsimony in explanations. Items 25 and 26 load negatively on this component so perhaps in some way the ability of scientists to form cogent parsimonious explanations is negatively affected by the quality of the data.

Component 7. Component 7 includes Items 6, 15, and 35. This component expresses the inability of scientists to be completely objective in spite of values such as complete disclosure, skepticism, and empiricism endorsed by most scientists.

Component 8. Items 2 and 12 load onto this component. Scientific methods can be utilized by people in many settings to gain a more complete understanding of various phenomena.

Component 9. This component consists of Items 9 and 37. These items are difficult to relate to each other. Item 9 states that scientists are creative and imaginative while Item 37 expresses the cyclical nature of the development of scientific knowledge.

Component 10. The tenth component includes Items 5 and 13. Both of these items express that theories are frequently validated through attempts to disprove them.

Component 11. Items 18 and 34 comprise component 11. Again, these items are difficult to draw meaning from. Item 18 expresses the inability of scientists to be creative, and Item 34 speaks of the need for critical peer review.

While some of the components group items together in meaningful ways, others are more obscure. In part, this emphasizes the complex overlapping of concepts that seems to exist in any description of the NOS. Interpreting the components is also affected by the low internal consistency reliability estimates. Items may also be grouping together in odd ways because the responses made by students reflect lack of instruction in the NOS.

Response Category Functioning

The category functioning statistics for all eight components exhibit behavior close to that predicted by the Rasch model. However, Table 10 reveals that some response categories do not function well for some subscales. In particular category 2 (*probably false*) does not appear to function as well as was hoped. Students either could not or did not attempt to distinguish between *probably false* as compared to *definitely false.* The preponderance of positively worded items as well as lack of student effort on the questionnaire could be explanations for this. For the time being, until further data can be collected, I recommend that the response options remain unchanged.

Potential Scale Usefulness for Science Teachers

This scale was developed as a more direct way to assess what secondary students understand about the nature of scientific inquiry and knowledge than the current assessment used by the state of Utah as described in Chapter 1 of this report. By separating nature of science questions from any subject-specific knowledge it was hoped that teachers might gain an understanding of what their students believe about aspects of the NOS which are essential to scientific work in any discipline. The issues regarding teaching and assessing knowledge of the NOS raised in this report are not unique to Utah.

This questionnaire may provide useful insights if administered to students throughout the United States.

Before using this questionnaire to guide instruction, the items identified for revision and deletion will be closely examined and altered accordingly. The questionnaire could then be re-administered and the responses re-examined for further improvements to the items. Ideally, the subscale structure might become clearer and perhaps more meaningful descriptors of the nature of science might be obtained from future administrations of the questionnaire.

In my classroom, I hope to be able to use this questionnaire at the beginning of the year in order to tailor instructional opportunities throughout the year to help present a correct understanding of science and scientific inquiry to my students so that they might be able to develop healthy understandings of the NOS. The questionnaire will be re-administered at the end of the year to gauge the effectiveness of instructional activities that have been employed. On a larger scale this questionnaire might be utilized by an entire science department, the schools within a district, or even statewide to assess students’ progress in developing correct knowledge about scientific inquiry as they work through a standard progression of topics and curricula throughout high school. To track individual students, identifying information would need to be included. For teachers in the classroom or in a department, this should pose no problem.

This questionnaire might also be used by university faculty to assess what in-service and pre-service teachers of science understand about the nature of science. As Lederman et al. (2002) point out, teachers can only teach the nature of science effectively to their own level of understanding. Experience throughout this project has highlighted that different levels of understanding about essential characteristics of science exist among scholars, philosophers, and teachers of science. It may be that pre-service and in-service teachers need more direct experience and instruction into the very aspects of science that this questionnaire was designed to assess.

In summary, I hope that this questionnaire can be an effective tool to help guide and focus instruction so that students will become thoughtful and knowledgeable consumers of scientific claims.

Recommendations

Recommendations for further work relating to this project are as follows:

1. A major issue hampering efforts to assess student knowledge of the NOS is the lack of explicit instruction regarding the NOS to students. It is likely that students do not receive direct, explicit instruction in the NOS because their teachers have never been explicitly taught about the NOS themselves. Institutions responsible for the preparation of science teacher candidates need to develop courses and coursework specifically designed to teach potential science teachers about the nature of science through experiential activities.
2. In addition to training provided to pre-service teachers, school districts and universities need to cooperatively implement in-service teacher training opportunities specifically designed to enhance science teachers' knowledge of the nature of science.
3. The indicators of ILO 6 do not completely reflect important aspects of the NOS present in the literature. Indicators similar to those presented at the bottom of Table 6 should be included in future versions of the Utah Core Curriculum in

Science.

4. Before the CSQ is re-administered to students, items need to be written to enhance the coverage of the components identified in Table 8 that have insufficient items.
5. The CSQ should be administered only under conditions where students are expected to take it seriously.
6. Instructions to teachers for administering the CSQ need to be standardized. Teachers administering the questionnaire should receive explanations of the relevance of responses to the questionnaire to their classes.
7. Motivation to respond honestly to the questions might be assessed directly by including a question at the beginning of the CSQ asking students to indicate their level of interest in science.
8. A differential item functioning analysis should be performed to determine if the items of the CSQ operate differently for students who have had three or more science classes as compared to those who have only completed one or two science classes.

References

Abd-El-Khalick, F. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education, 22,* 665-701.

Alwin, D. F. & Krosnick, J. A. (1991). The reliability of survey attitude measurement. *Sociological Methods and Research, 20*, 139-181.

American Association for the Advancement of Science. (1990). *Science for all Americans.* New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press.

Aikenhead, G. S., Fleming, R. W., & Ryan, A. G. (1987). High-school graduates' beliefs about science-technology-society: I. Methods and issues in monitoring students' views. *Science Education*, *71*, 145-161.

Anderson, L. W., & Bourke, S. F. (2000). *Assessing affective characteristics in the schools.* Mahwah, NJ: Lawrence Erlbaum Associates.

Chen, S. (2006). Development of an instrument to assess views on the nature of science and attitudes towards teaching science. *Science Education, 90,* 803-819.

Clough, M. P. (1998). Integrating the nature of science with student teaching: Rationales and strategies. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 197-208). The Netherlands: Kluwer Academic Publishers.

Cooley, W. & Klopfer, L. (1961). *Test on understanding science, Form W*. Princeton, NJ: Educational Testing Service

Dawis, R. V. (2000). Scale construction and psychometric considerations. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 65-94). San Diego, CA: Academic Press.

Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*, *25*, 283-289.

Good, R. C. (2000). Cautionary notes on assessment of understanding science concepts and nature of science. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 343-354). San Diego, CA: Academic Press.

Hogan, K. (2000). Exploring a process view of students' knowledge about the nature of science. *Science Education*, *84*, 51-70.

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

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire (VNOS): Towards valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*, 497.

Lederman, N. G., Wade, P., & Bell, R. L. (1998). Assessing understanding of the nature of science: A historical perspective. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 331-350). The Netherlands: Kluwer Academic Publishers.

Linacre J.M., & Wright B.D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, *8*, 360.

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51,* 493-504.

Martin-Diaz, M. J. (2006). Educational background, teaching experience and teachers' views on the inclusion of nature of science in the science curriculum. *International Journal of Science Education, 28,* 1161-1180.

McComas, W. F. (1998). The principal elements of the nature of science: Dispelling the myths. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 53-70). The Netherlands: Kluwer Academic Publishers.

McComas, W. F., Clough, M. P., & Almazroa, H. (1998). The role and character of the nature of science in science education. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 3-39). The Netherlands: Kluwer Academic Publishers.

McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standards documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41-52). The Netherlands: Kluwer Academic Publishers.

Munby, H. (1982). The impropriety of "panel of judges" validation in science attitude scales: A research comment. *Journal of Research in Science Teaching*, *19*, 617-619.

Munby, H. (1997). Issues of validity in science attitude measurement. *Journal of Research in Science Teaching*, *34*, 337-341.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and* *applications*. Thousand Oaks, CA: Sage.

Niaz, M. (2001). Understanding nature of science as progressive transitions in heuristic principles. *Science Education*, *85,* 684-690.

Nielson, E. R., (2002). *Investigating the effects of combining positively and negatively oriented items on the dimensionality of Likert scales.* Unpublished doctoral dissertation, Brigham Young University.

Rubba, P. A., & Andersen, H. O. (1978). Development of an instrument to assess secondary school science students' understanding of the nature of scientific knowledge. *Science Education*, *62,* 449-458.

Scharmann, L. C., & Smith, M. U. (2001). Further thoughts on defining versus describing the nature of science: A response to Niaz. *Science Education, 85,* 691-693.

Schulz, E. M., & Sun, A. (2001, April). *Identifying undifferentiating response sets and assessing their effects on the measurement of items*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Smith, M. U., & Scharmann, L. C. (1999). Defining versus describing the nature of science: A pragmatic analysis for classroom teachers and science educators. *Science Education*, *83,* 493-509.

Tedman, D. K., & Keeves, J. P. (2001). The development of scales to measure students', teachers' and scientists' views on STS. *International Education Journal, 2*, 20-48.

Utah State Office of Education (2002). *Utah Science Core Curriculum: Physics.* Retrieved January 5, 2005, from http://www.usoe.k12.ut.us/curr/science/core/ physics/docs /PHYSICSCORE03.DOC.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341-351.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Appendix A

First Pilot Version of the *Characteristics of Science Questionnaire*

Instructions:

What is science? What do scientists do as they seek new knowledge? What are the essential characteristics of the kinds of knowledge science and scientists are capable of generating?

Each statement below makes a claim about some essential characteristic of the kinds of work that scientists do and/or the kinds of knowledge they gain from their work. The claims may or may not be accurate. Carefully read each statement. Decide whether or not you believe the claim is correct and how strongly you feel about it. For each statement, select one of the following response options to indicate how strongly you agree or disagree with the claim being made.

Please be as honest and thoughtful as you can be about your answers. There is no ‘right’ or ‘wrong’ answer to any of the statements, it is just what *you* think or feel about it.

If you do not understand what a word means, you may raise your hand and ask your teacher.

Response options:

A. Strongly Agree
B. Agree
C. Slightly Agree
D. Slightly Disagree
E. Disagree
F. Strongly Disagree

There are 46 questions total.

1. Scientists strive to be unbiased when they analyze research results.
2. Scientific research methods can be used to make valid judgments about moral issues.
3. There are no limits to the types of questions that can be asked and answered by science.
4. Scientists value peer scrutiny so research is fully disclosed.
5. Scientists' curiosity about the natural world leads them to continually ask "What if…?" and "So what…?" questions.
6. Scientific methods focus on natural events which can be observed and/or measured.
7. Scientists form theories to link concepts, principles, hypotheses, and observations in a logical way.
8. The process of developing scientific knowledge is a repeated cycle of asking questions, collecting evidence, and forming conclusions.

9. Scientists describe their methods with enough detail for others to repeat because they expect other scientists to duplicate their work.
10. Science is essentially theoretical.
11. Conclusions drawn by previous scientists often form the foundation for new research.
12. Scientists use established rules of evidence to evaluate their own research findings and the findings of other researchers.
13. Scientists strive to be unbiased when they report research findings.
14. Scientific principles and laws are always tentative and subject to change as new evidence is available.
15. Scientists attempt to describe their findings with enough detail that other researchers can judge the adequacy of their procedures.
16. Scientists strive to develop theories and hypotheses that reflect the observable natural world.
17. Flaws in the arguments and explanations of previous research often lead scientists to plan and conduct new research.
18. Scientists value honesty in reporting the methods and procedures they use in their research.
19. Scientists seek the simplest explanations of data with the greatest ability to make accurate predictions.
20. Scientists present their findings as logical arguments.
21. Scientific hypotheses can be tested in an unbiased manner.
22. Once a scientific theory has been established, scientists stop further investigation.

23. Science is an attempt to explain observations of natural events.
24. Productive scientists have to be creative and imaginative.
25. A necessary part of science involves validating claims made by scientists by replicating research results many times.
26. Successful scientists habitually question what others accept as established knowledge.
27. When two theories explain the same data equally well, scientists prefer the simpler explanation.
28. Scientists value openness and honesty in reporting their own research and expect other scientists to do the same.
29. Scientists generate research questions and hypotheses from the theories they accept as true.
30. Scientists assume that the natural world operates according to constant rules that can be observed.
31. By critically analyzing previous research, scientists generate new questions and hypotheses to investigate.
32. Scientists value methods of observation that produce unbiased, replicable data.
33. Scientists frequently argue with other scientists about the assumptions, methods, findings and conclusions of their work.
34. The development of scientific knowledge is a cumulative process.
35. Scientific reasoning often involves creative leaps of imagination.
36. Scientists rarely use findings from previous investigations as they plan new research.

37. Scientific conclusions can never be completely proven because they are based on an inductive process.
38. Scientists use existing theories and explanations as a starting point for new research.
39. Scientists expect other scientists to generally accept research methods, evidence, and conclusions without question.
40. Scientists form theories to describe observations of the natural world.
41. The same step-by-step procedure commonly known as the 'Scientific Method' is used by scientists in all fields of research.
42. Theories do not allow scientists to make predictions of new events.
43. No matter how hard the scientist tries, observation methods and results are always tainted by their personal beliefs.
44. Interpretations of scientific evidence are presented as logical arguments in written reports; therefore, scientists strive to make the logic of their reasoning as clear as possible.
45. Science is a process of creating theories and then trying to prove them wrong.

46. How many science classes have you taken starting in 9th grade until now?

    Count each one separately; do not count a repeated class twice. Count a class and its follow-up AP course separately (example: count Chemistry and AP Chemistry as two classes). If you are not sure what to count, ask your teacher.

Appendix B

Revised Pilot Version of the

Characteristics of Science Questionnaire

Instructions:

This questionnaire is designed to assess your beliefs about the characteristics of science. Carefully read each statement. Decide how certain you are that the idea expressed is true or false. Choose an answer from one of the following categories:

A. Definitely True
B. Probably True
C. Probably False
D. Definitely False

Select the answer that best describes your belief about each statement. Completely fill in the circle on the answer sheet that corresponds to your answer. Do not skip any of the items.

Use a number 2 pencil only. If you change your mind about an answer, erase the previous mark completely before filling in the new answer. If you do not understand what a word means, raise your hand and ask your teacher.

1. Scientific knowledge is always based on observable evidence.
2. Science is a way of knowing that is often used by people other than scientists.
3. Scientific knowledge is always tentative and subject to change as new evidence becomes available.
4. The Scientific Method is a procedure that all true scientists use in the same manner.
5. Scientific inquiry is a process of testing theories by trying to disprove them.
6. Scientists can never be completely objective in the way they collect and interpret data.
7. Scientists frequently review and evaluate research performed by other scientists.
8. Absolute truth can never be determined by the methods of science.
9. Successful scientists are generally creative and imaginative persons.
10. Scientists often try to show that a theory is credible by demonstrating that it cannot be disproved.
11. When reporting results from their research, scientists have a responsibility to explain their thinking clearly enough that others can evaluate their reasoning.
12. Scientific inquiry is a continuing cycle of questioning ideas that are currently accepted as correct in hopes of gaining a more complete understanding.
13. The process of creating theories and then trying to disprove them is an essential part of science.
14. Successful scientists devote part of their time to evaluating previous research in order to plan new and better studies.

15. Scientific inquiry is characterized by a set of values that most scientists endorse (such as objectivity, skepticism, evidence that is observable, and honesty in reporting findings).
16. The results of a single scientific study can provide sufficient evidence to establish scientific truth.
17. When preparing written reports of their research, scientists attempt to describe their findings with enough detail that other researchers can judge the adequacy of their methods.
18. Scientific inquiry is always influenced by the values of the researcher.
19. In order to be labeled “scientific,” the findings of a research study must be based on observable evidence.
20. Scientists as a group are unwilling to accept new knowledge claims made by individual researchers until those claims have been confirmed by other researchers working independently.
21. Scientific research provides an objective way to answer questions about the benefits and risks of using modern technology.
22. The meaning of data collected by scientists is often subject to debate and interpretation by other scientists who represent different points of view.
23. Scientists should be both open-minded and skeptical when they evaluate the research conducted by other scientists.
24. Once a scientific truth has been established, there is no reason to question it further.
25. The validity of the conclusions reached by a scientist depends upon the quality of both the evidence collected and the reasoning used in drawing conclusions.

26. Scientists construct and use theories to explain recurring patterns they observe in the evidence they collect.
27. Since scientific inquiry is objective, it is not influenced by the values or aspirations of the researcher.
28. Ordinary people who are not trained scientists can use scientific rules of evidence to evaluate research claims that they encounter in advertisements and news reports.
29. Given a choice between two explanations that both explain their research findings, scientists generally prefer the simpler of the two explanations.
30. The scientific research process often raises ethical issues and moral questions that cannot be answered solely by the methods of science.
31. Science can never conclusively prove that an idea is true.
32. Science is based on the assumption that natural laws operate the same today as they have in the past.
33. Science is based on the assumption that natural laws will continue to operate the same way in the future as they do now.
34. A necessary part of the scientific process involves researchers critically reviewing the work completed by their peers.
35. Biology, Chemistry, Physics, Geology, and most other sciences share common standards for evaluating the knowledge claims made by scientists.
36. Since scientists strive to be objective, the inquiry methods they use and their interpretations of the results they obtain are not influenced by their experience or expectations.

37. Scientific research is a cyclical process of trying to create new knowledge by (a) formulating testable questions, (b) collecting evidence that bears on those questions, (c) drawing conclusions, and (d) formulating new, more refined questions.
38. Researchers from different scientific disciplines generally benefit more by working together on a single project than by working separately in their own area of expertise.

39. Since the beginning of ninth grade, how many semesters of science have you completed? Do *not* count any science classes in which you are currently enrolled, but have not yet finished.

    A None

    B 1 to 2 semesters

    C 3 to 4 semesters

    D 5 to 6 semesters

    E 7 or more semesters