Teacher Characteristics and Student Learning in Secondary Science

by

Eleanor Warfield Close

Dissertation

Presented to the Faculty of the

Graduate School of Education at

Seattle Pacific University

In Partial Fulfillment of the Requirements for the

Doctor of Education Degree

Seattle Pacific University

January 2009

Teacher Characteristics and Student Learning in Secondary Science

by

Eleanor Warfield Close

A dissertation submitted in partial fulfillment

of the requirement of the degree of

Doctor of Education

Seattle Pacific University

2009

Approved by ____

(Arthur K. Ellis, Ed.D., Chairperson of the Dissertation Committee)

____

(Rick Eigenbrood, Ph.D., Committee Member)

____

(Stamatis Vokos, Ph.D., Committee Member)

Program Authorized to Offer Degree ______________________________

Date ____

____

(Rick Eigenbrood, Ph.D., Dean, School of Education)

Copyright Page

Acknowledgements

The work of this dissertation would not have been possible without the support and help of many people. I wish especially to thank the following people:

Arthur Ellis, who offered steady encouragement and concise and insightful feedback;

Rick Eigenbrood, who generously shared his statistical expertise and answered innumerable detailed SPSS queries;

Stamatis Vokos, who asked questions and always kept the big picture in view;

My colleagues in the SPU Physics Department, who encouraged me to begin the doctoral program five years ago, and who have supported me in the task of juggling watermelons;

The faculty and staff of the SPU School of Education, who have gracefully accepted and supported me in my dual role as student and colleague;

Pam Kraus and Angie DiLoreto, who helped with study design and key background information;

Jennifer McKinney, who provided writing accountability, empathy, and regular infusions of chocolate;

My fellow doctoral students of Cohort 12, who shared countless hours of stories, study, and companionship;

My family, especially my husband Hunter Close, without whose ferocious and steadfast support this would have been inconceivable, and whose love and friendship are the joy of my life.

Dedication

To Hunter

and Hazel

and the little orcadillos

Table of Contents

List of Tables

List of Figures

Abstract

Teacher Characteristics and Student Learning in Secondary Science

Eleanor Close

Chairperson of the Dissertation Committee:

Dr. Arthur K. Ellis, School of Education

The question of how best to prepare and support K-12 science teachers for reformed teaching is a critical and unresolved issue. As described in the research review in chapter 2, many intermediate steps have been examined and documented; however, the link between teacher characteristics and student learning in science is not well studied. This study contributes to the knowledge base for the design of effective professional development for teachers of science.

This study examined relationships between teacher characteristics and student learning gains in secondary science. Participants in the study consisted of teachers (N = 15) and students (N = 1,250) in 8th grade public school science classrooms in a large school district in western Washington. Two measures were used to quantify student learning: student scores on the science portion of the Washington Assessment of Student Learning (WASL), and student pretest-posttest gain on researcher-designed content tests. Two instruments were used to collect information about teacher characteristics: the Epistemological Beliefs Assessment for Physical Science (EBAPS), and a Teacher Questionnaire.

A hierarchical multiple regression analysis was conducted to predict student gain scores from student WASL scores and teacher characteristics. The results of Step 1 of

this analysis indicated that WASL score accounted for a significant amount of the gain variability, $R^2 = .08$, $F(1, 640) = 52.04$, $p < .001$. Step 2 of the analysis indicated that teacher-related measures accounted for a small but statistically significant proportion of the student gain after controlling for student WASL score, R^2 change = .04, $F(3, 637) = 9.09$, $p < .001$. The full model accounted for approximately 11% of the variance in the criterion, $R^2 = .11$, adjusted $R^2 = .11$.

Consistent with previous research, relationships between student learning and individual teacher characteristics were mixed. Both teacher years of experience and teacher combined coursework were negatively correlated with student gain score ($r = -.16$, $p < .001$, and $r = -.21$, $p < .001$, respectively). Teacher EBAPS score was positively correlated with student gain score, $r = .12$, $p < .001$.

A second hierarchical multiple regression was conducted to examine relationships between teacher EBAPS score (criterion variable) and teacher years of experience and combined coursework (predictor variables). No significant correlation was found between the predictors and the criterion; neither the Step 1 regression model nor the full regression model was statistically significant.

Chapter 1: Introduction

Purpose of the Study

The purpose of this study is to examine relationships between teacher characteristics and student learning in 8th grade public school science classrooms in western Washington. Two measures are used to quantify student learning: student scores on the science portion of the Washington Assessment of Student Learning (WASL), and student gain from pretest to posttest on researcher-designed content tests administered in the science classes. Two instruments are used to collect information about teacher characteristics: the Epistemological Beliefs Assessment for Physical Science (EBAPS), and a Teacher Questionnaire incorporating questions about teacher coursework background, professional development, and school and administrative support.

Background

Americans have been concerned for some time about the level of understanding of science attained by students in our nation's schools, perhaps with special alarm beginning with Sputnik. The most recent set of reforms in science instruction has been in response to a report published in 1983, entitled *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education). This report, commissioned by the U. S. Secretary of Education, concluded (with some dramatic flair) that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (National Commission on Excellence in Education, 1983, p. 5). More recently, the report *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (Committee on Prospering in the Global Economy of the 21st Century &

Committee on Science, Engineering, and Public Policy, 2007) suggested "vastly improving" (p. 5) K-12 science and mathematics education as the first of four recommendations for ensuring America's continued economic viability. Science literacy for all Americans remains as yet an unmet goal.

Following the publication of *A Nation at Risk*, a number of reform efforts were begun. In the area of science literacy, the most influential publications resulting from these efforts on a national level have been those of Project 2061, a project of the American Association for the Advancement of Science (AAAS), and the *National Science Education Standards* (1996), published by the National Research Council (NRC).

Project 2061 is a "long-range, multi-phase effort designed to help the nation achieve scientific literacy" (AAAS, 1989, p. 11). As part of this effort, AAAS has produced a series of publications with the goal of reforming science education in the United States in such a way that all students attain literacy in science, mathematics, and technology. Each publication is the product of a lengthy process of discussion, debate, and revision, beginning with panels of experts in science and education, reviewed by the AAAS's National Council on Science and Technology Education, and approved by the AAAS Board of Directors. The first publication, *Science for All Americans* (AAAS, 1989), contains a set of recommendations on "what understandings and ways of thinking are essential for all citizens in a scientifically literate society" (p. 11), as well as chapters on effective learning and teaching and reforming education.

The content recommendations in *Science for All Americans* (AAAS, 1989) were intended to represent threshold levels for all students graduating from high school; the next Project 2061 publication, *Benchmarks for Science Literacy* (AAAS, 1993), contains recommendations (“benchmarks”) for what students should know and be able to do at particular grade levels. The *Benchmarks* were designed as a guide to the progress of students toward science literacy. The third Project 2061 publication, *Blueprints for Reform* (AAAS, 1998), defined organizational structures related to the education system and addresses what kinds of systemic changes are necessary to attain the goal of science literacy for all students. One of the structures addressed by the *Blueprints* is teacher education: the authors recognized that the reforms in science teaching advocated for the sake of science literacy require parallel reforms in the education of science teachers.

One section of the chapter on effective learning and teaching in *Science for All Americans* (AAAS, 1989) contains a list of aspects of good science teaching, including “start with questions about nature,” “engage students actively,” “concentrate on the collection and use of evidence,” “use a team approach,” and “do not separate knowing from finding out” (pp. 147-148). The following section described scientific values that should be reflected in science teaching, such as rewarding creativity, encouraging healthy questioning, and avoiding dogmatism (pp. 149-150). The *Blueprints for Reform* (AAAS, 1998) examined the role of teacher education and professional development in equipping science teachers with the knowledge and skills to be able to teach in ways consistent with the guidelines expressed in *Science for All Americans*. Among the skills and knowledge required are basic science literacy for the teachers themselves (p. 191), the ability to recognize and address students’ preconceptions about science (p. 190), and some personal experience, during their own science education, with teaching styles that support the reform effort (p. 196).

In a similar time frame to the AAAS's Project 2061, the NRC constructed and published a second set of national recommendations for science education, the *National Science Education Standards* (NRC, 1996). At the request of the National Science Teachers Association, and funded by the Department of Education and the National Science Foundation, the NRC established the National Committee on Science Education Standards and Assessment, which formed working groups to construct the first draft of the *Standards*. The committee then solicited input from scientists, science teachers, and others interested in science education, which was incorporated into the published document. The purpose of the NRC in developing the *Standards* was to promote scientific literacy, and to define a vision of science education reform that makes scientific literacy attainable for all students. Scientific literacy is defined as "the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity" (p. 22), and also includes specific content as set forth in the *Standards*.

The *Standards* (NRC, 1996), like the *Benchmarks* (AAAS, 1993), is a document consisting of sets of standards that define scientific literacy and describe steps toward attaining it. The standards for science content outline "what students should know, understand, and be able to do" in science during the progression from kindergarten through 12^{th} grade (NRC, 1996, p. 6). In addition, the *Standards* gives teachers guidance on how to provide an environment in which students can attain scientific literacy (similar to recommendations found in *Science for All Americans*, AAAS, 1989). These include standards for science content, science teaching, assessment, and professional development for science teachers. The *Standards* also includes program standards and system standards that address the responsibilities of policy makers and the wider community in helping make scientific literacy possible for all students.

Among the standards for science teaching are that teachers should “plan an inquiry-based science program for their students” (p. 30), “guide and facilitate learning” (p. 32), “engage in ongoing assessment of their teaching and of student learning” (p. 37), “design and manage learning environments that provide students with the time, space, and resources needed for learning science” (p. 43), and “develop communities of science learners that reflect the intellectual rigor of scientific inquiry and the attitudes and social values conducive to science learning” (pp. 45-46). In part, to meet these standards, teachers need to model scientific inquiry skills, monitor the understanding of individual students, encourage and facilitate discussions about scientific ideas, and engage students in significant decision-making about the course of their investigations.

The standards for professional development for teachers of science are designed to enable and support teachers in meeting the standards for teaching science (NRC, 1996). Similar in content to the recommendations found in *Blueprints for Reform* (AAAS, 1998), these standards state that professional development for science teachers must include “learning essential science content through the perspectives and methods of inquiry” (p. 59), “integrating knowledge of science, learning, pedagogy, and students, … [and] applying that knowledge to science teaching” (p. 62), and “building understanding and ability for lifelong learning” (p. 68). The point is made that in order to teach science effectively, teachers must have knowledge of both science content and pedagogy, and also of what Shulman (1986) named “pedagogical content knowledge”: knowledge of what students are likely to know and understand (or misunderstand) in specific content areas, and content-specific strategies to help them develop their understanding.

The documents discussed above give an overview of what we want our science teachers to be able to do, and outline some of the kinds of education and professional development that might be needed to enable teachers to teach in this way. This leads to a number of related questions. What do we know about the effects of professional development on teachers' understanding of science and on their ways of teaching science? Does professional development in science affect teachers' behavior in the science classroom? Does it affect their students' learning? How are teacher characteristics, such as coursework background and epistemological beliefs, related to student learning? How does teacher preparation, both preservice and in-service, alter relevant teacher characteristics?

Little empirical research has been done that directly relates teacher characteristics to student learning, and those studies that have been done have found mixed or inconclusive results (Greenwald, Hedges, & Laine, 1996; Monk, 1994; Wayne & Youngs, 2003; Wilson, Floden, & Ferrini-Mundy, 2002). Other research examined intermediate steps. For example, researchers in physics education have found that instructional method is related to student learning: research-based, inquiry-oriented instruction leads to greater improvement in conceptual understanding than do lecture-based instruction and traditional problem-solving activities (Hake, 1998; Kim & Pak, 2002; Mazur, 1997; McDermott & Shaffer, 1992; McDermott, Shaffer, & Constantinou, 2000). In addition, instructional method may negatively influence student beliefs about the nature of science and of teaching and learning about science (AAAS, 1998; Grossman

& Stodolsky, 1995; McDermott et al., 2000). Deep content knowledge on the part of teachers is a necessary, but not sufficient, condition for the kind of teaching advocated by the current reform movement (Cohen, 1990; Shulman, 1987).

Research on students' and teachers' conceptions about the nature of science has similarly mixed results. Research in the 1960s and 70s suggested that on average, both teachers and students had poor understanding of the nature of science (Lederman, 1992). Despite efforts in the post-Sputnik years to develop pre-college science curricula consistent with the nature of science, students who participated in these curricula did not develop the desired understanding of the nature of science (Abd-El-Khalick, Bell, & Lederman, 1998). Teachers' conceptions of the nature of science did not correlate with their college science credits, college grade-point average, or other relevant academic variables (Lederman, 1992). More recent research has focused on improving preservice and in-service teachers' conceptions, with some success (Akerson, Abd-El-Khalick, & Lederman, 2000; Akerson, Morrison, & McDuffie, 2006). While a sophisticated understanding of the nature of science was not found to guarantee teaching methods consistent with the nature of science (Lederman, 1992; Mellado, 1997; Schwartz & Lederman, 2002; Tobin & McRobbie, 1997), a naïve understanding appears to preclude such teaching methods (Cohen, 1990; Grossman & Stodalsky, 1995; Yerrick, Parke, & Nugent, 1997).

This study is designed to contribute to the research base relating teacher characteristics to student learning, taking into account the research summarized above.

Research Questions

This study will focus on the following research questions:

1. Is there a relationship between student gain on the researcher-designed content tests and student scores on the Science WASL?
2. Beyond any relationship between students' gain scores on the researcher-designed content tests and their scores on the Science WASL, is there a relationship between students' gain on the researcher-designed content tests and teachers' epistemological beliefs, coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, coursework in science content or science education, or years of experience?
3. Is there a relationship between teachers' epistemological beliefs and their years of teaching experience?
4. Beyond any relationship between teachers' epistemological beliefs and their years of teaching experience, is there a relationship between teachers' epistemological beliefs and their coursework in science content or science education, in inquiry-oriented teaching strategies, or in formative assessment and evaluation of students' prior knowledge?

Significance of the Study

The documents defining the current science education reform movement identify the kinds of science content understanding and conceptions about the nature of science students should develop during their K-12 education. Little empirical evidence exists, however, about the kinds of teacher education and professional development that lead to the desired student learning. More research is needed in this area to inform the design of both preservice and in-service teacher education. This study was informed by existing research relating teacher characteristics to student learning, as well as research examining other relevant teaching and learning relationships. The results of this analysis contribute to the knowledge base for the design of effective professional development for teachers of science.

Outline of Work

The researcher-designed content pretest and posttest were developed in collaboration with physics faculty at Seattle Pacific University and researchers at FACET Innovations, LLC, a science education research and development company. The content included in these tests was designed to assess student knowledge related to a particular unit of study common to all participating schools. The tests were administered by collaborating science teachers to 8th grade science students at the beginning and end of the relevant unit of study during the 2005-2006 academic year. Student scores on the science portion of the 2006 WASL were obtained from the Washington Office of Superintendent of Public Instruction. The science WASL is administered to all 8th grade students in Washington State in April or May.

The teacher questionnaire was also developed in collaboration with physics faculty at Seattle Pacific University and researchers at FACET Innovations, LLC. The questionnaire probed teacher characteristics such as coursework background, professional development, and perceived school and administrative support. The design of the questionnaire was informed by previous research on relationships between teacher characteristics and student learning. Collaborating science teachers completed the questionnaire online between spring and autumn of 2007. The EBAPS survey was administered to most collaborating science teachers in March of 2007, in pencil-and-paper format. A few teachers were not present at the professional development session at which the EBAPS was administered; these teachers completed the survey in an online format in summer or fall of 2007.

Descriptive and inferential statistics were computed using the SPSS software package (version 15.0). Hierarchical multiple regression was used to look for relationships among the data, as described in the research questions.

Chapter 2 contains a review of the literature relevant to this study. In chapter 3, the research methods used are described in more detail. Chapter 4 contains the results of the data analysis, and chapter 5 includes discussion of the results of the analysis and suggested directions for future research.

Chapter 2: Review of Literature

Relevant Philosophy and Theory

The vision for science education presented in the *National Science Education Standards* (National Research Council [NRC], 1996) and the Project 2061 documents (American Association for the Advancement of Science [AAAS], 1989, 1993, 1998) is based on a constructivist understanding of teaching and learning. Constructivism has as its central premise the idea that "the learner constructs all knowledge from previously acquired knowledge, personally, socially, or in combination" (Ellis, 2001, p. 130); science education therefore must provide students with personal experiences on which basis they can construct scientific knowledge. The most prominent early work in American constructivist educational theory is that of John Dewey. Jerome Bruner contributed to constructivist theory on the mind and learning, and applied the ideas of cultural psychology to the realm of education. Lee Shulman, president of the Carnegie Foundation for the Advancement of Teaching and Professor of Education Emeritus at Stanford University, developed a constructivist theoretical framework for content knowledge in teaching.

Constructivism

In his book *Experience and Education* (1938), Dewey described education as a collection of experiences on the part of the student, and described the kind of learning environment conducive to desirable educational experiences. According to Dewey, "experiences in order to be educative must lead out into an expanding world of subject-matter, a subject-matter of facts or information and of ideas" (1938, p. 87). In *The School and Society* (1915), Dewey described such an experience for an elementary student as one in which “the child [is] not simply doing things, but getting also the *idea* of what he does; getting from the start some intellectual conception that enters into his practice and enriches it” (p. 76). The ideas developed are, in turn, applied in future experiences, both enriching the experiences and expanding the ideas.

The role of teachers in this learning environment is to arrange fruitful experiences for their students. In order to accomplish this, teachers must be familiar with the needs and abilities of their students, and must be able to determine what experiences will meet those needs and extend those abilities (Dewey, 1938, p. 58). In Dewey’s theoretical framework of teaching and learning, the scientific method is the key both to determining what educational experiences will be most useful, and to the usefulness of those experiences for students. Dewey advocated “systematic utilization of scientific method as the pattern and ideal of intelligent exploration and exploitation of the potentialities inherent in experience” (1938, p. 86). In education, as in science, ideas must be carefully crafted and rigorously tested through observation. The test of an idea lies in the observed outcome of the implementation of the idea. Accumulation of careful observations and reflection on them makes it possible to accept, extend, or revise ideas in education as in science: “to reflect is to look back over what has been done so as to extract the net meanings which are the capital stock for intelligent dealing with further experiences” (Dewey, 1938, p. 87). For students, this means reflection on their experiences in the classroom, and further development of the ability to understand the information and ideas of the subjects being studied; for the teacher, this means reflection on the intellectual growth of students produced by particular sets of classroom experiences, and further development of the ability to select fruitful experiences for students consistent with their observed needs and abilities.

This philosophy of education makes the separation of teaching method from taught content an artificial divide. In *Democracy and Education: An Introduction to the Philosophy of Education* (1916), Dewey described method as “the effective direction of subject matter to desired results” (p. 194). This effective direction requires teachers to understand both the content to be learned and the students who will be learning it, and to be aware of what other teachers have done that has lead to other students successfully developing their understanding of that content: “study of the operations and results of those in the past who have greatly succeeded is essential” (Dewey, 1916, p. 200). Teaching methods must remain connected to what is taught and informed by reflection on the results in the classroom: “Never is method something outside of the material” (Dewey, 1916, p. 194).

Bruner applied ideas from cultural psychology to the realm of education in a way that is consistent with Dewey’s theories. In his book *The Culture of Education* (1996), Bruner listed nine tenets for an approach to education guided by cultural psychology. Among these are the *perspectival tenet* (p. 13), the *constructivism tenet* (p. 19), the *interactional tenet* (p. 20), and the *instrumentalism tenet* (p. 25). Briefly summarized, the *perspectival tenet* is that meaning is influenced by perspective: a person’s understanding and interpretation of an event or a piece of information will depend on the person’s culture and experience. The *constructivism tenet* is that what we think of as “reality” is constructed through a process of meaning making; a process which, as discussed in the *perspectival tenet*, is shaped by the perspective of the person making the meaning. Both these tenets relate to Bruner's idea expressed in earlier works that thought, experience, and meaning-making all are shaped by language: "the very language one speaks conditions the style and structure of thought and experience" (1962, p. 116; see also 1966, pp. 102-112).

The *interactional tenet* is that learning is an interactive process. Bruner argued that the kinds of interaction between teacher and student that take place in the traditional "transmission model" of instruction are perhaps the least conducive to the development of knowledge and skill, because they are one-way interactions: the teacher tells or shows things to the students, who are expected to absorb what is told or shown to them (1996, p. 20-21). A classroom more consistent with the cultural-psychological approach could be described as "a subcommunity of mutual learners, with the teacher orchestrating the proceedings" (Bruner, 1996, p. 21). As with the ideal learning environment described by Dewey in *Experience and education* (1938), students in this subcommunity would work together to build on prior experiences, receive guidance from their teacher, reflect on meaning and develop understanding. In earlier works, Bruner described the process of building on experience in the context of content learning, arguing that "the organizing ideas of any body of knowledge are inventions for rendering experience economical and connected" (1962, p. 120) and that meaningful learning in any subject area can be experienced by students of all ages (see, e.g., 1961, pp. 12-13 and pp. 33-54; 1966, p. 35).

The *instrumentalism tenet* is that education, of any form and in every culture, is instrumental in shaping the lives of those who experience it. This closely echoes Dewey, who wrote that "wholly independent of desire or intent, every experience lives on in further experiences" (1938, p. 27). In this way, education shapes not only our understanding of the world we live in but also our ways of functioning within our culture. Schools, therefore, are not only a place where students learn facts and ideas, but also where they become socialized into the larger culture. This means that “the chief subject matter of school, viewed culturally, is school itself” (Bruner, 1996, p. 28). The meaning-making done by students about the “subject matter” of school is inextricably interwoven with the meaning they make of their culture and life experience.

In the area of science specifically, Bruner distinguished between “dead” or “finished” science and “the lively processes of science making” (1996, pp. 126-127). The process of learning science, from this perspective, should not be primarily becoming familiar with an account of the accomplishments of past scientists. Rather, it should be a process of learning to make sense of experiences in a systematic way, which can only be done by the learner herself; “all one can do for a learner en route to her forming a view of her own is to aid and abet her on her own voyage” (Bruner, 1996, p. 115). Bruner expressed this somewhat more emphatically in *The Process of Education* (1962): “The schoolboy learning physics *is* a physicist, and it is easier for him to learn physics behaving like a physicist than doing something else” (p. 14). Learning science in this way requires that the learner have experiences similar to those of a practicing scientist: she must learn by doing, reflecting, and constructing understanding from and about her experience.

Shulman’s Three Kinds of Content Knowledge

In his article “Those Who Understand: Knowledge Growth in Teaching” (1986), Lee Shulman suggested a theoretical framework for content knowledge in teaching. He divided content knowledge into three categories: subject matter content knowledge;

pedagogical content knowledge; and curricular knowledge (Shulman, 1986, p. 9). Subject matter content knowledge is knowledge of the field to be taught; not only of the facts and concepts within the field, but also of how we know them, the relationships between them, and why we consider them worth knowing. Shulman elaborated on this category in "Knowledge and Teaching: Foundations of the New Reform" (1987):

> A teacher… must understand the structures of subject matter, the principles of conceptual organization, and the principles of inquiry that help answer two kinds of questions in each field: What are the important ideas and skills in this domain? and How are new ideas added and deficient ones dropped by those who produce knowledge in this area? That is, what are the rules and procedures of good scholarship or inquiry? (p. 9)

Teachers serve as the primary source of content understanding for students, and students learn both ideas and values in and about the content from teachers, whether explicitly or implicitly. Thus, in addition to the key concepts and results in the content realm, teachers must also be familiar with how ideas and concepts are developed as part of their subject matter content knowledge.

Shulman's definition of content knowledge incorporates how scientists construct knowledge; the definition of pedagogical content knowledge centers on how students construct their own understanding of this scientific content knowledge. Shulman defined pedagogical content knowledge as "the particular form of content knowledge that embodies the aspects of content most germane to its teachability" (Shulman, 1986, p. 9). This includes knowledge of multiple representations of important ideas within the field, as well as explanations, examples, and analogies that are most productive in teaching. It also encompasses knowledge of difficulties frequently encountered by students when learning topics in the field, commonly held preconceptions, and strategies that can be used to help students reorganize and refine their understanding. As described by both Dewey and Bruner, the meaning made by a student in any situation is shaped by that student's culture and previous experiences; therefore, teachers need knowledge of the ideas students have developed and how those ideas will influence the students' construction of new, related ideas.

Curricular knowledge consists of knowledge of curricular materials available for teaching specific content, as well as understanding of what curriculum students are likely to have encountered in previous classes or will likely encounter in future classes in related content areas. Shulman made a comparison between physicians and teachers: just as we expect a physician to be aware of the full range of treatment options for a particular ailment, we should expect and prepare teachers to be aware of the full range of curricular options for a particular topic. This is consistent with Dewey's description of the role of the teacher as arranger of fruitful experiences for students, based on knowledge of the students' needs and abilities and of the kind of experience that will meet their needs and extend their abilities (Dewey, 1938, p. 58).

Nature of Science

One component of scientific literacy, as defined in both *Science for All Americans* (AAAS, 1989) and *National Science Education Standards* (NRC, 1996), is an understanding of the nature of science (NOS). The entire first chapter of *Science for All Americans* is devoted to the NOS; in the *National Science Education Standards*, the history and nature of science is one of eight science content strands. In fact, understanding of the NOS has been an instructional goal in American education since at least 1907 (Lederman, 1992, p. 332), and a wide variety of research studies have been done documenting the teaching and learning of the NOS in K-12 classrooms and in teacher preparation programs.

Science for All Americans (AAAS, 1989) described the NOS in three main categories. The first category, “the scientific worldview,” includes basic beliefs and attitudes of scientists, including the idea that the world is understandable; and that scientific ideas are tentative and subject to change, but many ideas become more stable and less tentative over time as evidence is accumulated. The second category, “scientific inquiry,” describes elements of the practice of science common across scientific disciplines, such as the central role of observations in evaluating scientific claims; the combination of logic and creativity used in advancing science; and the role of debate and consensus in building scientific knowledge. The third category, “the scientific enterprise,” characterizes science as a social activity, influenced by wider social values and carried out in a variety of settings.

The description of the NOS in the *National Science Education Standards* (NRC, 1996) is similar to the description in *Science for All Americans* (AAAS, 1989), emphasizing an understanding of science as a human endeavor and the primacy of data and observational evidence in the evaluation of scientific explanations. Like *Science for All Americans,* the *National Science Education Standards* differentiated between established ideas in science, which are less tentative and less likely to change in the future, and newer, more currently active areas of science, in which ideas are not yet well established and it is normal for scientists to disagree. Particularly for younger students, the *National Science Education Standards* recommend an emphasis on experiencing scientific investigations and thinking about explanations, rather than memorization of vocabulary and information.

In his review of the research, Lederman (1992) stated, “although the ‘nature of science’ has been defined in numerous ways, it most commonly refers to the values and assumptions inherent to the development of scientific knowledge” (p. 331). Some specific aspects that are generally accepted by the science education community as important for scientific literacy are: scientific knowledge is tentative, rather than certain or absolute; subjective (influenced by the theories of the researcher), rather than objective; and creatively constructed based on empirical data, rather than discovered (AAAS, 1989; Akerson, Abd-El-Khalick, & Lederman, 2000; Elby & Hammer, 2001). These are consistent with the descriptions of the NOS found in *Science for All Americans* (AAAS, 1989) and the *National Science Education Standards* (NRC, 1996), though the emphasis found in many research studies is somewhat different, as discussed below.

Some theory and research relevant to understanding of the NOS is found in the philosophical area of epistemology. Epistemology is “an area of philosophy concerned with the nature and justification of knowledge” (Hofer & Pintrich, 1997, p. 88). While epistemological research and theory is not specific to the area of science, it applies to the nature and justification of knowledge in science as well as in other fields. In their critical review of epistemological research, Hofer and Pintrich synthesized the major ideas from different research programs into a consensus set of four epistemological dimensions in two categories. In the category “Nature of Knowledge,” the authors described the dimensions “certainty of knowledge” and “simplicity of knowledge” (pp. 119-120). In the dimension “certainty of knowledge,” the authors described the naïve view that knowledge is certain and absolute, and the sophisticated view that knowledge is “tentative and evolving” (p. 120). In the “simplicity of knowledge” dimension, the authors described the naïve view that knowledge is a body of discrete, concrete facts, and the sophisticated view that knowledge is composed of interrelated, contextual concepts. In the category “Nature of Knowing,” the authors described the dimensions “source of knowledge” and “justification for knowing” (p. 120). “Source of knowledge” describes whether a person understands knowledge as being transmitted by authority or personally constructed; “justification of knowledge” includes evaluation of evidence and methods for judging knowledge claims (p. 120).

While Hofer and Pintrich (1997) described the dimensions above as aspects of personal epistemological theories, these constructs are consistent with the NOS as described above. One important consideration in comparing research in the areas of epistemology and the NOS is the distinction between personal and public epistemologies (Lising & Elby, 2005). The research synthesized by Hofer and Pintrich addresses personal epistemology, that is, a person’s ideas about his or her own knowledge and learning. The NOS encompasses both personal and public epistemologies: a person’s set of ideas about the NOS in the context of the scientific community is an inherently public epistemology; however, a person’s set of ideas about his or her own knowledge and learning in the area of science is a personal epistemological stance. Lising and Elby (2005) argued that a student’s public and personal epistemologies need not be consistent. For example, a student may believe that active scientists construct coherent scientific knowledge, but may at the same time see her own scientific knowledge as lacking coherence and perhaps lacking the possibility of coherence (p. 373).

Elby and Hammer (2001) challenged the characterization of naïve and sophisticated views described by Hofer and Pintrich (1997), and at the same time identified a weakness of the majority of research on student and teacher ideas about the NOS. The challenge consisted of two charges: first, the consensus view of desirable epistemological beliefs fails to distinguish between “correct” beliefs and “productive” ones; and second, that the widely accepted characterization of a sophisticated epistemology fails to take context into account. For example, as described in *Science for All Americans* (AAAS, 1989) and the *National Science Education Standards* (NRC, 1996), some ideas in science are more tentative than others; therefore, when considering whether a scientific idea is tentative or certain, the specific scientific idea under consideration is important contextual information. However, most research on student and teacher ideas about the NOS, including Hofer and Pintrich’s synthesis, has considered the broadly generalized tentative view to be the sophisticated understanding. In addition to being generalized to the point of being incorrect, Elby and Hammer argued that the broad statement of scientific knowledge as tentative was also likely to be less productive for many students than the idea that scientific knowledge is absolute: a student who thinks of science as tentative might be less inclined to struggle with apparently contradictory ideas than a student who holds the view that scientific knowledge is absolutely true. A truly sophisticated epistemology, according to Elby and Hammer, must take into account the context (both the field within science and the specific scientific idea within that field) and the intended use of the knowledge or idea under consideration (p. 565).

While the details of what defines the NOS may be subject to debate, as described above, all definitions of the NOS are explicitly constructivist. In addition to being clearly consistent with constructivist theory as described by both Dewey and Bruner, the definition of scientific knowledge as tentative and subjective is consistent with Bruner's perspectival tenet: all knowledge construction is shaped by the perspective of the person doing the constructing. Shulman's definition of subject matter content knowledge includes the principles of inquiry in the content domain, and therefore encompasses the constructivist aspect of the NOS as well as the tentative and subjective elements: if knowledge is constructed (not discovered) and ideas are dropped as well as added, then the ideas are shaped by the scientists constructing the knowledge (and are therefore subjective), and the knowledge constructed is inherently tentative and subject to future rejection in favor of new ideas. Since the NOS is part of the content knowledge of science, pedagogical content knowledge for science teachers must include understanding of how students construct their understanding of the NOS, as well as knowledge of what kinds of difficulties are common to students learning about the NOS.

One strand of research related to the NOS concerns epistemology, the set of ideas about knowledge and learning. As described above, individuals have both personal and public epistemologies (Lising & Elby, 2005); public epistemology includes beliefs about knowledge construction by the scientific community, whereas personal epistemology includes beliefs about construction or acquisition of personal knowledge. While most definitions of the NOS describe a public epistemology concerning the construction of knowledge by the scientific community, personal epistemologies influence how students construct their own scientific knowledge (Lising & Elby, 2005). Understanding of the NOS is therefore both a component of scientific literacy and an influence on the process of constructing scientific understanding.

Critical Open Questions

The goal of the science education reform movement is science literacy for all Americans, and the most critical component in attaining this goal is improved K-12 science teaching. To be consistent with the reform movement, the science experienced by K-12 students must be more constructivist, more student-centered, and more consistent with the nature of science than is the case in most classrooms today. Though this kind of science teaching has existed in some classrooms for a long time, and has been advocated by science education reformers for decades, it has not often been successfully implemented on a large scale (AAAS, 1998, p. 190). Therefore, preparation for preservice teachers, and continuing professional development for in-service teachers, must equip them for reform-based teaching and support them in its implementation. For practicing teachers who have not experienced reformed science instruction, either as teachers or as learners, this is especially important since implementing reformed science teaching will require a shift away from familiar and comfortable instructional methods and possibly a shift in understanding about the nature of science as well. In order to teach in a way consistent with the nature of science, teachers must be comfortable with the idea that "science is not just a body of knowledge, but a paradigm through which to see the world" (AAAS, 1998, p. 202); for teachers who learned science primarily as a body of facts, making the transition to reformed-based science teaching will be particularly difficult.

The Project 2061 document *Blueprints for Reform* (AAAS, 1998) stated that "high quality professional development for K-12 faculty and administrators is essential to science education reform" (p. 223). As described chapter 1, both the *Blueprints for Reform* and the *National Science Education Standards* (NRC, 1996) give general guidelines for what kind of professional development is needed to support teachers in implementing reformed science instruction. However, as noted in the *Blueprints,* more research is needed in order to provide evidence about the effectiveness of specific professional development strategies in influencing teachers to adopt reform-based science teaching methods (p. 82). Carrying out this kind of research requires the development of an effective method of measuring the extent to which classroom teaching is in fact consistent with reform-based principles, which has been a challenging task:

> Despite the plethora of literature advocating a shift in teaching and learning of science and mathematics toward student-centeredness, the development of sensitive evaluation frameworks and data collection instruments appropriately aligned to these efforts has been both difficult (National Institute for Science Education, 1999) and controversial (Linn, 2000). (Sawada et al., 2002, p. 245).

In addition to evaluating the classroom practice of teachers, research on the effects of science education reform efforts requires evaluation of student learning. Given the long-term goal of science literacy for all high school graduates, the final test of any reform effort must be its impact on the understanding of science students. As part of a report on the development of a reform-based classroom observation protocol, Sawada et al. stated that such evidence is not yet convincing, and "in the absence of strong quantitative evidence, debate over the impact of reform rages unabated" (2002, p. 245). This is consistent with the consensus represented in *Blueprints for Reform* (AAAS, 1998): while there is general agreement that teachers play a critical role in helping students attain science literacy, there is no similar consensus about "what it means to be an effective teacher" and what kind of preparation and professional development contributes to effective teaching (p. 82). One chapter in the *Blueprints* suggested a research agenda to support reform, including questions about teacher preparation and effectiveness:

> How and where do effective K-12 science teachers learn to teach? How much of their effectiveness can be attributed to the influence of college science teachers? How much and what aspects of their effectiveness can be explained by their knowledge and understanding of science? (p. 88).

A 2002 report on preservice teacher preparation commissioned by the Office of Educational Research and Improvement and the U. S. Department of Education again described a shortage of rigorous research on the effects of either subject matter knowledge preparation or pedagogical preparation on teacher knowledge and student learning (Wilson, Floden, & Ferrini-Mundy, 2002).

In summary, the issue of how best to prepare and support K-12 science teachers for reformed teaching is a critical and unresolved question. While some principles of effective teaching are described in the science education reform literature, there is a scarcity of rigorous research on the effects of teacher preparation and professional development on teaching practice and student learning. As described in the research review below, many intermediate steps have been examined and documented; however, the critical link between teacher characteristics and student learning in science is not well studied.

Research Context

Research on Relationships between Teacher Characteristics and Student Learning

A large body of research has been done attempting to identify the best ways to assess teacher quality and to influence student learning. Many of these studies have found that teacher characteristics are important variables affecting student achievement (e.g., Fetler, 1999; Greenwald, Hedges, & Laine, 1996; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Wayne & Youngs, 2003; Wright, Horn, & Sanders, 1997). However, pinpointing the particular teacher characteristics that matter to student learning is difficult. For example, a large-scale study of students in grades 3-5 using the Tennessee Value Added Assessment System found that the teacher is the most important factor affecting student learning gains in all content areas (Wright et al., 1997); however, the study did not include any data that could illuminate what made a teacher effective. A similar study in Texas examined student performance in reading and mathematics in grades 3-7, with similar results: teachers were found to have a large influence on student reading and math achievement, but specific teacher characteristics examined in the study did not explain the variation in teacher effects (Rivkin et al., 2005). Using data from a three-year randomized experimental study (the Tennessee STAR project), another analysis also found that teachers have a large impact on student learning (Nye et al., 2004), and also found that the specific teacher characteristics included in the study (experience level and degree level) were not very successful in predicting teacher effectiveness. The randomized assignment of both student and teachers to classrooms makes this a particularly convincing result supporting the importance of teacher characteristics for student learning, but the study does not provide much indication of what characteristics cause the teacher effects.

One common research method for examining relationships between teacher characteristics and student achievement is the use of education production functions. Also known as process-product research, studies of this type are based on the economics-influenced model of schools as producers of student achievement, and use regression analysis to estimate the size of relationships between school resources (including teachers) and achievement while statistically controlling for possibly confounding background characteristics. In their 1996 meta-analysis, Greenwald, Hedges, and Laine aggregated data from 60 education production function studies to look for statistically significant relationships between student achievement and a variety of school inputs. All studies included in the analysis used the results of student scores on standardized achievement tests as the outcome measure; in most cases these were tests of reading, mathematics, or verbal ability. Other selection criteria for studies to be included were that the data were published in a refereed journal or book, the data originated in the United States, the level of aggregation of the data was at the school district level or below, the model in the study either controlled for socioeconomic characteristics or was longitudinal or quasi-longitudinal in design, and the data were stochastically independent from other data included in the universe of studies (in order to avoid bias due, e.g., to the inclusion of multiple test scores for the same students) (pp. 364-365).

The seven input variables selected by Greenwald et al. (1996) for analysis fell into three categories: expenditure variables, including per-pupil expenditures and teacher salaries; teacher background characteristics, including teacher ability, teacher education, and teacher experience; and size, including school size and class size (p. 365). The results of the analysis of teacher background variables showed that both advanced education (such as a master's degree) and "high ability" in teachers were strongly and significantly positively correlated with student achievement. The methods used to measure teacher ability were not specified by Greenwald et al., but examination of some of the studies used in the meta-analysis reveals a variety of measurements used to quantify teacher ability, including average score on the National Teacher Evaluation test given by the Educational Testing Service to new teachers (for data aggregated at the school district level) (Strauss & Sawyer, 1986); teacher score on a vocabulary test (Levin, 1970, p. 68; Michelson, 1970, p.123); and average selectivity (aggregated at the school level) of the institutions at which teachers received their bachelor's degrees, based on Barron's college rankings (Ehrenberg & Brewer, 1994, p. 4). Because these measurements do not fall into the realm of content-area professional development, they will not be explored further here.

The results of the Greenwald et al. (1996) meta-analysis suggest that, on average, advanced education for teachers improves student achievement. However, the analysis did not differentiate between content education and other forms of advanced education for teachers, leaving open the question of what kinds of professional development for teachers are most helpful for student learning. One difficulty in interpreting the body of research on teacher characteristics, including the Greenwald et al. meta-analysis, is the wide variety in the kind of data used to characterize teachers. For example, some of the studies included in the Greenwald et al. meta-analysis treated advanced education for teachers as a dichotomous variable (master's degree or no master's degree), while others treated it as a continuous variable (Greenwald et al., 1996, p. 367). In one study, the authors divided teacher education into five categories: BA, BA + 15, BA + 30, MA, and MA + 15 (Lewis & Ouellette, 1979, p. 3). Most studies used only degree status or number of credits of post-degree coursework (e.g., Lewis & Ouellette, 1979; Rivkin et al., 2005), while some include information about the content area of the coursework or degree in relation to the subjects being taught (e.g., Monk, 1994; see also Wayne & Youngs, 2003).

While the meta-analysis by Greenwald et al. (1996) showed a positive correlation between teacher advanced education and student achievement, many studies have found no relationship. In their systematic review of research on teacher characteristics and student learning gains, Wayne and Youngs (2003) found that most studies of the effects of teacher degrees and coursework found indeterminate results; among studies that did find statistically significant relationships, some were found to be positive and some negative (p. 101). Rather than statistically combining the results of the reviewed studies, Wayne and Youngs grouped studies according to the teacher characteristic addressed. Within each group, the authors then described the strengths and weakness of each study, analyzed specific details that influenced what inferences could be drawn in each case, and then arrived at a joint interpretation of the group. In the group of studies addressing the effects of teacher degrees and coursework, the authors noted that most studies (most of which found indeterminate results) used data sets containing teacher degree level only.

However, more recent studies have included information about the content area of the degree and its relationship to the subject being taught. Wayne and Youngs noted that the results of these more recent studies suggest that the indeterminate or conflicting results of earlier studies are at least partly explained by the absence of subject-related information (p. 101).

The clearest examples of the importance of coursework or degree subject area are in mathematics. Several studies of the relationship of teacher characteristics to mathematics learning gains in high school students found different results when teacher degree level was considered with and without content-specific information (Wayne & Youngs, 2003). When teacher degree level (without the degree subject specified) was examined, these studies found that students whose teachers had master's degrees achieved the same learning gains as students whose teachers did not have master's degrees. However, when the subject area of the master's degree was included in the analysis, differences did emerge: students of teachers with a master's degree in mathematics learned more than students whose teachers had either no master's degree or a master's in a subject other than mathematics (p. 102). Similar results were found for teachers' bachelor's degree (mathematics vs. other subjects), and another study found a positive correlation between the total count of mathematics courses taken by teachers and their high school students' mathematics learning (p. 103). A separate study of the mathematics achievement of high school students in California found consistent results for mathematics certification: a strong negative correlation was found between student test scores and the percentage of mathematics faculty who were emergency-certified (i.e. who lacked standard mathematics certification) (Fetler, 1999).

In contrast to the results Wayne and Youngs (2003) reported for mathematics, the results of studies on relationships between teacher preparation in science and student science learning gains are less clear. Wayne and Youngs reported only two studies with statistically significant results in this area, and the results are not consistent. In a study using student and teacher data from the National Educational Longitudinal Study of 1988 (NELS:88), Goldhaber and Brewer (1997) compared the science learning gains of students whose teachers had bachelor's degree in science to those whose teachers had bachelor's degrees in other subject areas. The researchers found that students of teachers with a bachelor's degree in science had higher learning gains. The difference between the groups was found to be small but statistically significant.

In the second study examined by Wayne and Youngs (2003), Monk and King (1994) used data from the Longitudinal Study of American Youth (LSAY), which provided more specific information about teacher preparation including counts of undergraduate and graduate coursework in mathematics, physical science, and life science. Monk and King categorized teacher preparation in physical science and life science separately, and looked for relationships between each area of preparation and student learning gains for two grade levels (10^{th} and 11^{th}). Wayne and Youngs reported that the only statistically significant finding of this study was a small negative correlation between teacher physical science preparation and junior student learning gains (significant at the .05 level), a finding that seems to contradict the notion that higher levels of teacher subject preparation lead to larger student learning gains. Monk and King also reported small positive relationships between teacher physical science preparation and sophomore student learning gains, significant at the .10 level, and noted that an earlier analysis (Monk, 1994) looked in more detail at the relationships between teacher preparation and student learning.

The earlier analysis by Monk (1994) was not included by Wayne and Youngs (2003) in their literature review because it lacked controls for students' socioeconomic status. As Monk and King (1994) noted, however, the results of this analysis are consistent and more detailed, and are therefore worth examining here. In this study, which was included in the Greenwald et al. (1996) meta-analysis, Monk (1994) used data from the LSAY to look for relationships between student achievement in math and science and 15 different teacher education characteristics. Teacher education was characterized by the following: number of courses in life science, at undergraduate and graduate levels; number of courses (undergraduate and graduate) in physical science (chemistry, physics, or earth science); number of courses (undergraduate and graduate) in mathematics; number of science education courses (undergraduate and graduate); number of mathematics education courses (undergraduate and graduate); whether an undergraduate science major was obtained; whether an undergraduate math major was obtained; whether a master's degree was obtained; whether additional credits were earned after obtaining a master's degree; and years of teacher experience (Monk, 1994, pp. 127-128). All teacher education data were self-reported (Monk, 1994, p. 129). Student achievement and achievement gain in mathematics and science were assessed using a composite of scores on test items developed by the National Assessment of Educational Progress, from achievement tests administered in the fall of their sophomore, junior, and senior years in high school (Monk, 1994, p. 126). Data were gathered from a nation-wide stratified random sample of 2,829 students and their teachers of science ($N = 483$) and mathematics ($N = 608$) (Monk, 1994, pp. 126-128).

The results of Monk's (1994) analysis were complex. In most cases, more undergraduate mathematics or science courses taken by teachers correlated with a small increase in student mathematics or science performance (pp. 130-137). In mathematics, the effect size was greater for juniors than for sophomores, and for juniors it diminished when the teacher had five or more courses (p. 130). In science, undergraduate life science preparation had a negative effect on student science achievement for juniors; the relationship had a threshold effect, however, with the negative relationship holding only for teachers with fewer than six courses in life science (p. 136). Graduate coursework in mathematics correlated with improved student mathematics achievement for sophomores (p. 132); for juniors, graduate coursework in life science showed a positive effect on science achievement, but graduate coursework in physical science showed a negative effect (p. 136).

Undergraduate science and mathematics education courses contributed more to student learning than did undergraduate science and mathematics courses (Monk, 1994, p. 130). A teacher having majored in mathematics had no apparent effect on student mathematics performance at either the sophomore or junior level (p. 132); a teacher having majored in science had a positive effect on student performance in science (p. 137). A teacher's degree level was found to have insignificant or negative correlation with student achievement gains in mathematics (p. 132) and in science (p. 137). Monk also noted that the correlation between degree level and number of courses in the content area was “quite modest” (p. 132), suggesting that in cases when content knowledge is of primary interest, teacher degree level is not a very useful measure.

In addition to the direct results discussed above, Monk (1994) found interaction effects between teacher preparation and student ability. In mathematics, for example, the regression model based on teacher characteristics explained 60% of variation in performance of high-scoring sophomores, but only 10% in performance of low-scoring sophomores (p. 132), indicating an interaction effect between teacher preparation and initial student performance. In science, Monk found interaction effects between undergraduate coursework and subject taught. For example, the relationship between teacher undergraduate coursework in physical science and student achievement was positive for sophomores enrolled in life science classes, but negative for sophomores enrolled in physical science (p. 141).

The complexity of the relationships in Monk’s (1994) study suggests that it is a waste of time to look for simple relationships between, for example, level of teacher coursework and overall student achievement. Teacher degree level does not appear to be a useful subject of study in exploring ways to improve student performance. Content-area education classes, however, were positively related to student learning, with up to double the effect size of content courses (p. 137), suggesting that “a good grasp of one’s subject area is a necessary but not a sufficient condition for effective teaching” (Monk, 1994, p. 142). This appears to reinforce Shulman’s idea that pedagogical content knowledge, in addition to subject-area content knowledge, is necessary for effective teaching. It is also worth noting that while student achievement was measured using content-related multiple-choice assessments—a direct, if potential controversial, method of measuring student understanding—teacher understanding was not directly assessed. Self-reported counts of subject area coursework and content-related education coursework served as proxies for teacher content knowledge and teacher pedagogical content knowledge, respectively. Since the understanding gained by taking a course depends in complex ways on a wide variety of factors, as demonstrated by this study, course counts offer at best a rough estimate of teacher understanding. Relying on these measures places severe limits on our ability to meaningfully explore relationships between teacher characteristics and student progress.

A 2003 study by Adamson et al. looked more closely at the type of science course taken by preservice teachers in an effort to assess the impact of reforms in mathematics and science teaching at Arizona State University. The Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) is a National Science Foundation supported project with the goal of improving undergraduate instruction in mathematics and science, with a particular focus on courses taken by preservice teachers. The ACEPT Evaluation Facilitation Group developed an observation protocol to make a holistic measurement of the degree to which observed teaching of mathematics and science is "reformed." The group defined reformed teaching to be characterized by the constructivist, student-centered methods advocated by the NRC (e.g., 1996), AAAS Project 2061 (1989, 1993), and the National Council of Teachers of Mathematics (Sawada et al., 2002). The resulting Reformed Teaching Observation Protocol (RTOP) had very high interrater reliability for trained observers.

The researchers studied a sample of 28 secondary mathematics and science teachers and their students (Adamson et al., 2003). Data was collected on the undergraduate preparation of each teacher, and teachers were characterized by the number of ACEPT-influenced courses taken. ACEPT-influenced courses were defined as those taught by faculty who had participated in the ACEPT faculty workshops, in which they were introduced to reformed teaching methods and began the process of revising their course practices. Other research has found that ACEPT-influenced courses are characterized by higher RTOP scores and higher student learning gains than non-ACEPT courses (Sawada et al., 2002).

The RTOP was used to characterize the level of reform teaching by each secondary teacher. In addition, data was collected on the biology achievement of the students of 15 biology teachers (a subset of the secondary teacher sample) using a multiple-choice assessment designed to measure knowledge of biology concepts, scientific reasoning, and the nature of science. In comparing groups, the researchers controlled for student socioeconomic status and teacher years of experience, and also for section type (regular or honors) in the analysis of student biology achievement. Comparisons of ACEPT-influenced teachers (those who had taken one or more ACEPT-influenced courses as undergraduates) to non-ACEPT teachers showed that ACEPT-influenced teachers had significantly higher RTOP scores (Adamson et al., 2003). The researchers also compared teachers with one ACEPT-influenced coursed to those with two or more such courses, but the differences between these two groups were not statistically significant. Comparisons of student biology achievement showed that students of ACEPT-influenced teachers had significantly higher biology achievement scores.

In addition to providing evidence for the effectiveness of reformed teaching methods in science, this study (Adamson et al., 2003) is consistent with Monk's (1994) findings that course counts, even if differentiated by content area, do not yield much information about teacher effectiveness. If the teaching methods implemented by the instructors of preservice courses have a strong influence on the effectiveness of in-service teaching, as is suggested by this study, then a simple count of content coursework misses a critical variable: the degree of constructivist teaching practiced in each of the counted courses. Unfortunately, this level of course characterization is not available in large national datasets such as the LSAY and the NELS:88, and may in fact be impossible to obtain on a national scale.

In contrast with the large-scale quantitative research described above (e.g., Greenwald et al., 1996; Monk, 1994; Monk & Kings, 1994; Wayne & Youngs, 2003), Shulman in 2002 reflected on the last 30 years of teacher education research influenced by the Institute for Research on Teaching (IRT). The IRT research program was designed on the basis of a critique of process-product research methods (Shulman, 2002, p. 248); in the 1970s, Shulman and his colleagues at the IRT worked "to redirect the paradigms of teaching by shifting attention from behavior to thought; from observable performance to strategy and understanding; and from simple models of stimulus and response to more complex and subtle models involving context, content, and cognition" (Shulman, 2002, pp. 249-250). This research paradigm, implemented at the IRT and later continued at Stanford University and the Carnegie Foundation, led to significant advances in understanding of teacher knowledge, including the development of the concept of pedagogical content knowledge.

In moving away from the process-product research paradigm, however, Shulman and his colleagues rejected the use of standardized test scores as adequate measures of student understanding but failed to replace them with any other indicator of student learning. Thus, the research program in teacher understanding was disconnected from the study of student learning: “Baby was discarded along with bath water” (Shulman, 2002, p. 250). According to Shulman, only in the last few years (since 2000) have researchers in this paradigm begun to re-connect teacher characteristics with student learning outcomes (p. 251). This may account in part for the dearth of empirical evidence documenting relationships between teacher content understanding and student achievement. As described above, many studies have attempted to connect teacher characteristics to student learning, but few have included content-specific information (Wayne & Youngs, 2003; Wilson, Floden, & Ferrini-Mundy, 2002) and most have not been successful in finding meaningful relationships (Greenwald et al., 1996; Wayne & Youngs, 2003).

Around the time of Shulman’s reflections, Wilson, Floden, and Ferrini-Mundy (2002) were commissioned by the U. S. Department of Education to summarize existing empirical research on five key questions, including “What kind of subject matter preparation, and how much of it, do prospective teachers need? Are there differences by grade level or subject area?” (Wilson et al., 2002, p. 191). Their selection criteria for existing research included evidence of disciplined inquiry and analysis described in sufficient detail as to be replicable. In reviewing empirical, peer-reviewed research published in the United States in the last two decades, they found no studies, from any research methodology, that directly measured teacher understanding of content and related teacher understanding to student learning (p. 191). Even when teacher coursework or undergraduate major were accepted as proxies for content understanding, Wilson et al. found only seven studies that met their selection criteria (p. 191). Similarly, in their 2003 review of the literature addressing teacher characteristics and student learning gains, Wayne and Youngs found a total of only 21 publications meeting their selection criteria (which included data accounting for students' prior achievement and socioeconomic status), only two of which addressed content coursework or degrees for science teachers.

Research relating professional development for teachers and their students' learning is similarly scarce. In *How people learn: Bridging research and practice*, a report based on a synthesis of current research on human learning, Donovan, Bransford, and Pellegrino (1999) reported that not much empirical evidence existed regarding what kind and what amount of professional development is needed to change teacher performance and student achievement significantly (pp. 46-47). This report is consistent with that of Greenwald et al. (1996), who wrote that research previous to their meta-analysis did not show strong or consistent results about the relationship between resources (such as teacher education and ability) and student achievement (pp. 361-362).

Given the scarcity of relevant studies relating teacher understanding to student learning, it is worth considering studies documenting intermediate steps, such as relationships between teachers' coursework and their epistemological beliefs, content understanding, and choice of teaching methods.

Research on Relationships between Coursework and Teacher or Learner Characteristics

If we accept the premise that teacher understanding of content is an essential condition for student learning, as proposed in Science for All Americans (AAAS, 1989, p. 191), we should examine evidence exploring how much and what kind of content understanding teachers gain from various kinds of education or professional development.

In many introductory college-level classes the primary instructional method is lecture. Numerous research studies have found that lecture instruction leads to little improvement in conceptual understanding (see, e.g., Hake, 1998; McDermott & Shaffer, 1992; McDermott, Shaffer, & Constantinou, 2000). Coursework in the physical sciences also typically involves extensive practice in problem solving. However, increased facility in solving traditional numerical problems does not necessarily indicate increased conceptual understanding. Working with introductory physics students at Harvard University, Mazur (1997) found that after traditional instruction, most students were able to successfully solve traditional computational exam questions involving complex electrical circuits; however, the same students were for the most part unable to answer relatively simple conceptual questions about less complex electrical circuits. This suggests that numerical problem solving, while it may increase student familiarity with the mathematical formalism that represents physical relationships, does not typically improve student understanding of the conceptual structure of the relationships themselves; if conceptual understanding were similarly improved, students should be able to solve non-numerical conceptual problems as easily as numerical ones.

In a study of Korean university students, Kim and Pak (2002) found that after solving between 300 and 2900 (self-reported) traditional physics textbook problems, students still demonstrated difficulties with conceptual questions in basic Newtonian mechanics, and there was little correlation between conceptual understanding and number of problems solved. Kim and Pak measured conceptual understanding using the Force Concept Inventory (Hestenes, Wells, & Swackhammer, 1992), a non-numerical test of Newtonian mechanics concepts that uses everyday language rather than the formal structure common in traditional textbook problems. Since traditional science instruction (at both the undergraduate and graduate level) often focuses on more computational, less conceptual, treatment of science content, the results found by Mazur (1997) and Kim and Pak (2002) suggest that traditional science courses are poor preparation for science teaching.

Advanced coursework does not necessarily lead to deeper understanding of introductory-level topics, or of the structures and conceptual principles relating ideas within the subject area. Physics majors and graduate students in physics doctoral programs often retain misconceptions about introductory physics topics, or are unable to articulate the conceptual structure of introductory ideas. The Physics Education Group at the University of Washington collected vast quantities of data on student understanding of a wide variety of topics in introductory physics, including data from advanced undergraduate physics majors and graduate student Teaching Assistants in the physics doctoral program. These data consistently showed that advanced coursework usually does not result in deep conceptual understanding of introductory topics (e.g., McDermott & Shaffer, 1992; Wosilait, Heron, Shaffer, & McDermott, 1999). Since most K-12 science content falls into the category of “introductory” with respect to college or university courses, advanced-level content courses for teachers are not necessarily an effective way of improving teacher understanding of content relevant to their teaching.

Some colleges and universities now offer special content courses designed for teachers. As is emphasized in both the *National Science Education Standards* (NRC, 1996) and the *Blueprints for Reform* (AAAS, 1998), as well as in *How People Learn: Bridging Research and Practice* (Donovan, Bransford, & Pellegrino, 1999), teachers need to experience inquiry-based, active-engagement instruction in order to be able to implement these methods in their own classrooms; thus many teacher-oriented content courses are taught with little or no lecture, and are structured instead around guided inquiry and reflection. Emphasis on inquiry is not sufficient, however, to guarantee conceptual understanding of the content being taught. McDermott, Shaffer, and Constantinou (2000) reported a comparison of two groups of prospective elementary teachers enrolled in the teacher education program at the University of Cyprus. Teachers were randomly assigned to one of two instructional treatments (L. C. McDermott, personal communication, June 27, 2005). The first group used a Greek translation of *Physics by Inquiry* (McDermott & University of Washington Physics Education Group, 1996), a curriculum based on the findings of discipline-based education research. The second group learned the same content following constructivist educational methodology, but without incorporating the results of discipline-based education research (for example, specific student difficulties identified by research were not explicitly addressed). Two posttests were used to asses student understanding: a set of free-response questions that asked students to explain the reasoning used to determine their answers; and a multiple-choice conceptual test developed by a third party (McDermott et al., 2000, p. 415).

On both posttests, students in the first group had average scores two to four times greater than students in the second group. McDermott et al. (2000) concluded that students in the second group achieved conceptual understanding comparable to that seen in traditional lecture-based courses (see McDermott & Shaffer, 1992), while students in the first group demonstrated significantly higher levels of understanding. Some students in the first group had completed instruction on the content tested a year before the exams were administered; these students achieved the same posttest averages as those who completed instruction immediately before the exam, showing excellent retention (McDermott et al., 2000, p. 415). These results demonstrated that Shulman's ideas apply as much to teaching teachers as to teaching K-12 students: the instructor must be familiar not only with the content to be taught, and with good pedagogy, but also with how the learner is likely to interact with the content and what ideas the learner may bring to the content.

As stated above, in many introductory college-level classes the primary instructional method is lecture, and research suggests lecture instruction leads to little improvement in conceptual understanding. For future or current teachers, lecture-based instruction has additional disadvantages. Teachers tend to teach as they have been taught (AAAS, 1998; Lawson et al., 2002; McDermott et al., 2000); what they have learned through lecture, they are likely to have difficulty adapting to other forms of instruction. This suggests that inquiry-based, active-engagement instruction is needed for teachers not only for the sake of their own conceptual understanding of the science content, but for the development of their pedagogical content knowledge.

A study by the Arizona State University ACEPT group bears this out. The research team used the RTOP to assess the level of reformed teaching by first- through third-year middle and high school teachers who were recent graduates, some of whom had experience reformed teaching in ACEPT-influenced courses and some of whom had not (Judson & Sawada, 2001; Lawson et al., 2002). ACEPT-influenced courses were those in which the teaching methods were consistent with constructivist, reformed pedagogy. The results showed a statistically significant difference between the two groups: new teachers who had experienced constructivist, inquiry-based coursework during their teacher preparation program demonstrated teaching practice in their own classrooms that was more consistent with the science education reform movement. This finding supports the notion that the experience of constructing their own knowledge in this kind of classroom environment is an important component in enabling teachers to develop their understanding of the kinds of instruction that will help their students, in turn, to construct the desired scientific knowledge.

Prospective secondary teachers typically major or minor in the subject(s) they will be teaching. Courses within university departments shape students’ understanding of the discipline not only explicitly, through content development and discipline-specific ways of approaching questions, but also implicitly, through the organization of courses within the department and the way in which the department fits into the larger institution (Grossman & Stodolsky, 1995, pp. 8-9). For example, introductory mathematics and science courses are frequently designed to weed out the majority of students enrolled in them; this may shape students' beliefs about whether mathematics and science should be accessible to the average student (Grossman & Stodolsky, 1995, p. 9). Teachers whose beliefs are shaped by this kind of experience are likely to avoid instructional strategies that are based on the premise that all students can learn content at some meaningful level.

McDermott et al. (2000) noted that elementary and middle school teachers typically do not have the prerequisites for standard introductory science courses, especially in the physical sciences. The courses available to them in college science departments are usually survey courses, which cover a wide range of material with not much depth and very little conceptual grounding (p. 412). This type of course can reinforce the idea that science consists of a set of facts to memorize, rather than an active process that teachers and students can understand and in which they can participate. Thus, preservice teachers' understanding of the NOS is shaped to be consistent with the process of learning science in traditional survey courses.

Early research on understanding of the NOS focused on K-12 students and their teachers, and then on the coursework preparation of K-12 teachers and how that might influence their NOS understanding (Lederman, 1992). The earliest studies were designed to assess the general state of student understanding of the NOS. One of the first such studies, by Mead and Metraux (1957), was a qualitative assessment of high school students' ideas about science and scientists. Students participating in the study responded to one of three essay prompts. A representative random sample was then drawn from the 35,000 responses gathered, and the selected responses were analyzed for patterns. While the researchers were primarily concerned with students' ideas about science as a possible career path, their results shed some light as well on respondents' views of the NOS.

Among the patterns found were that the students thought of science as advanced by hard work alone, and not involving creativity or imagination; and they thought of the scientist as a person working alone for long periods of time before “he finds out something” (p. 388). These responses are consistent with the idea of science as a fixed body of facts that are discovered, rather than scientific knowledge as tentative and creatively constructed by scientists.

Unlike the study by Mead and Metraux (1957), most of the early research was quantitative in nature, using multiple-choice or short-answer surveys to characterize students’ and teachers’ conceptions of the NOS. The most commonly used instrument was the *Test On Understanding Science* (TOUS), a multiple-choice survey developed by Klopfer and Cooley in 1961 (Lederman, 1992, p. 333). In 1963, Miller used the TOUS to compare the understanding of science of 7^{th} through 12^{th} grade students in Iowa with each other and with the understanding of high school biology teachers. Miller found that while students in high school scored better on the TOUS than students in the 7^{th} and 8^{th} grades, only high-ability 11^{th} and 12^{th} graders had an average score above 50%. He also found that the average score of the high-ability 11^{th} and 12^{th} grade students was not statistically significantly different from the average score of a sample of 51 biology teachers from 20 different schools in Iowa, and that 68% of this high-ability student group had higher scores than 25% of the teachers sampled. Miller interpreted these results as a dramatic failing of the educational system, and called for reform in college teaching practices as well as science teacher preparation programs. While Miller’s methods of selection of subjects are not well described, his findings suggested a significant shortfall in understanding of the NOS, both for students and for many of their teachers.

Also using the TOUS, Mackay (1971) pretested and posttested over 1200 students in 7th through 10th grade in three randomly-selected high schools in the Melbourne, Australia metropolitan area. The test was administered at the beginning and end of the 1968 academic year. The students were separated into eight groups for analysis: each grade level (7th, 8th, 9th, and 10th) was separated into two groups (boys and girls). Only five of the eight groups made statistically significant gains in TOUS score from pretest to posttest; even among those groups with significant gains, the largest difference in mean score between the pretest and posttest is less than two points out of 60 possible.

Mackay also compared performance on specific items between students at the beginning of 7th grade and students at the end of 10th grade. All students were required to take science in 7th and 8th grades, whereas 9th and 10th grade students self-selected to take science; therefore, the average student taking science in 10th grade should be expected overall to be more successful in science than the average student taking science in 7th grade. Despite this difference in student populations, Mackay found that on over a third of the TOUS items, student performance did not improve between the 7th and 10th grades. Mackay concluded that the science instruction students experienced between 7th and 10th grade was deficient in the extent to which it addressed the NOS. A number of specific deficiencies in students' understanding of the NOS were also identified from the TOUS data, including understanding of the role of creativity in scientific research and the relationship between scientific experiments, models, theories and absolute truth. These results are consistent with the idea that most secondary science coursework does not significantly impact student understanding of the NOS.

Student ideas about the relationship between scientific research and absolute truth were also addressed by Horner and Rubba (1978). In their article “The myth of absolute truth,” the authors discussed possible causes of one of the findings of Rubba’s 1977 doctoral research: that 30 percent of science students at a high school in the American Midwest believed “that scientific research reveals incontrovertible, necessary, absolute truth” (p. 29). The authors noted that most textbooks represent scientific knowledge as discovered truth, and also that previous research suggested that science teachers may not have a sufficient understanding of the NOS teach science as tentative. In a subsequent article, Horner and Rubba described “The Laws are Mature Theories Fable” (1979), another commonly misunderstood aspect of the NOS.

As demonstrated in the examples described above, early research results suggested that on average students had poor understanding of the NOS, and also that typical coursework did not have much effect on student NOS conceptions. Based on the assumption that students’ conceptions were influenced by the understandings of their teachers, researchers undertook similar studies to explore the NOS understanding of teachers. Results of these studies indicated that on average, teachers’ conceptions of the NOS were also inadequate or naïve. For example, Miller’s 1963 study (which examined the NOS understanding of both students and teachers) indicated that many high school biology teachers knew less about the NOS than their students. Schmidt (1967) replicated Miller’s study and found similar results, though fewer students in Schmidt’s study scored higher than 25% and 50% of the teachers than did so in Miller’s study. Schmidt then administered the TOUS to 116 scientists in Iowa working in university and industry settings and found a mean score of 50.8, out of 60 possible points. Selection methods for the scientists are not described, nor are the number of science students in each group in the replication study. Schmidt concluded that the problem of low teacher NOS understanding described by Miller still existed; also that a lower score on the TOUS might be considered “respectable” than had previously been thought. These results suggest that the coursework experienced by teachers between the time they were secondary students and the time they began teaching science did not do much to improve their understanding of the NOS.

Carey and Stauss (1968) studied the NOS conceptions of 17 preservice secondary science teachers enrolled in a science methods class, using an open-ended essay prompt and the *Wisconsin Inventory of Science Processes* (WISP), an instrument consisting of 93 statements with which test-takers agree or disagree. Their analysis of the essay responses showed that fewer than half the preservice teachers described science as a body of knowledge, as methods and modes of inquiry, and as a human endeavor, though all but one participant used at least one of these descriptions. The responses also included statements that equated science with technology, with nature, or with scientists, as well as describing science as truth. Participants’ WISP scores averaged 68.2 at the beginning of the methods class and 72.4 at the end. While they did not state what a desirable WISP score would be, Carey and Stauss concluded that many prospective secondary science teachers did not possess an adequate understanding of the nature of science.

In 1970, Carey and Stauss published a similar study in which they used the WISP to pretest and posttest 31 in-service science teachers enrolled in an Academic Year Institute. The teachers’ experience ranged from 1 to 10 years, with a mean of 5.24 years. The authors concluded that, like the prospective teachers in the previous study (1968), the practicing teachers participating in this study did not have a sufficient level of NOS understanding.

In his 1967 study, Kimball developed an instrument (the *Nature of Science Survey*) and used it to assess the NOS understanding of science teachers and practicing scientists with similar academic backgrounds. One premise of this research study was that previous research on science teacher NOS understanding was biased because of the common practice of sampling from practicing science teachers, many of whom might be teaching outside of their content expertise. To criticize science teacher preparation programs based on the NOS understanding of teachers who did not actually prepare to teach science, Kimball argued, is an unfair conclusion. Instead, a comparison should be made between science majors who choose to become teachers and those who choose to become practicing scientists.

Kimball found that when academic background was controlled for, there was no significant difference in the NOS conceptions of science majors who became teachers and those who became practicing scientists. This does not mean, however, that Kimball found science teachers to have adequate conceptions of the NOS: neither group demonstrated strong NOS understanding, with mean scores for both groups around 35 out of 58 possible points on the survey. Also, comparisons of recently graduated groups to groups with graduation dates over a decade before showed no significant difference, suggesting that neither scientists nor science teachers improved their NOS conceptions over time. Based on these results, Kimball suggested that criticism over the NOS preparation of science teachers could be directed at undergraduate science programs, rather than (or in addition to) teacher preparation programs. This result is consistent with the notion that typical coursework for a science major does not lead students to develop sophisticated understanding of the NOS. A more recent study also supports this idea: preservice teachers who had completed or almost completed a major in biology were found to have very low scores on a NOS survey (Lawson et al., 2002).

Further research was undertaken to explore relationships between teachers' NOS understanding and their academic backgrounds. A series of studies done in the late 1960s and 1970s found that teachers' conceptions of the NOS (as measured by multiple-choice surveys) did not correlate with academic variables such as college grade-point average, college science credits, particular college science courses, high school science coursework, or mathematics grades (Lederman, 1992, p. 340). For example, in their 1968 and 1970 studies described above, Carey and Stauss collected data for each participant on a number of academic variables related to science preparation, including both hours of coursework and grade average in biological science, physical science, and mathematics. In both studies, they found little or no statistically significant correlation between WISP score and any of the academic variables. In a third, larger-scale study, Carey and Stauss (1969) looked for correlations between academic variables and WISP scores for a sample of 35 prospective secondary teachers and 221 prospective elementary teachers enrolled in methods classes at the University of Georgia over a one-year period. Again they found few correlations that were statistically significant, and the few correlations that were found to be significant had small correlation coefficients. The authors concluded that little or no relationship exists between academic variables and understanding of the NOS as measured by WISP.

In addition to looking at the relationship between science preparation and NOS understanding, Carey and Stauss (1968, 1970) used WISP pretest and posttest scores to examine whether a science education course with an emphasis on the nature of science could improve students' NOS conceptions. For both groups (17 prospective teachers in the 1968 study, 31 practicing teachers in the 1970 study), the average WISP posttest score was significantly greater than the average pretest score. Thus, the researchers concluded that science education coursework could improve teachers' NOS understanding.

Building on these results, Billeh and Hasan (1975) used the multiple-choice *Nature of Science Test*, previously constructed by Hasan, to pretest and posttest four groups of secondary science teachers participating in a four-week summer institute. Teachers were divided by interest and teaching assignment into chemistry, biology, physics, and physical science groups. All four groups received instruction in science teaching methods, participated in laboratory activities emphasizing guided discovery, and experienced enrichment materials showing scientists at work; in addition, all groups except the biology group received 12 50-minute lectures about the NOS. The authors found that all groups except the biology group had significant gains in their scores on the *Nature of Science Test* from pretest to posttest, confirming the results of Carey and Stauss (1968, 1970) that formal instruction in the NOS can improve teacher conceptions. This study had the additional result that those teachers participating in a methodology course that did *not* include formal NOS instruction (biology teachers) did *not* improve their understanding of NOS as measured by the *Nature of Science Test*. Teacher gain scores were also found to be unrelated to educational background (two-year or four-year degree), previous in-service training, or subject taught.

While Carey and Stauss (1968, 1970) and Billeh and Hasan (1975) found that targeted coursework could improve teacher NOS understanding, these and other studies suggested that teacher NOS conceptions did not seem to improve as a result of the process of teaching science. Carey and Stauss (1970) found no relationship between teaching experience and WISP score in their sample of 31 practicing teachers with between 1 and 10 years of experience. Billeh and Hasan grouped their sample of 141 practicing teachers, all of whom had been teaching for between one and nine years, into those with two or fewer years of experience and those with more than two years of experience, and found no difference in NOS conceptions between the two groups. Kimball (1967) found the same result in his sample of 78 preservice and in-service teachers with science majors.

One assumption embedded in early NOS research and not recognized until the 1990s was the existence of a causal relationship between participation in the activities of science and developing understanding of the NOS. Science curricula developed in the 1960s and 1970s, in the post-Sputnik wave of science education reform, emphasized hands-on, inquiry-based learning, with the underlying assumption that by participating in these activities students would come to understand not only the content but the NOS as well. Research studies of the effects of these curricula have not supported this assumption; the hands-on, inquiry-based curricula of the 1960s and 1970s were not found to lead students to develop the desired conceptions of the NOS (Abd-El-Khalick, Bell, & Lederman, 1998, p. 419). This helps explain the lack of correlation between coursework in science and understanding of the NOS for preservice and in-service teachers: even a science major including substantial laboratory research may not influence students' ideas about the NOS. Beyond the experiences included in an undergraduate science major, even working as a practicing scientist may not influence a person's views of the NOS: as described above, Kimball's 1967 study found that neither scientists nor science teachers improved their NOS conceptions over time.

A large body of recent research has focused on the question of how to improve preservice and in-service teachers' conceptions of the NOS. Influenced by previous research showing lack of correlation between science coursework and improved NOS conceptions, much of this research has been done in the context of science methods courses for preservice teachers (e.g., Akerson, Abd-El-Khalick, & Lederman, 2000; Shapiro, 1996). Because previous research suggested that science instruction in which the NOS remained implicit (such as the curricula of the 1960s and 1970s reform movement) did not lead to improvement in participants' understanding of the NOS, and some evidence suggested that explicit NOS instruction could be effective (e.g., Billeh & Hasan, 1975; Carey & Stauss, 1968, 1970), some of these studies focused on courses with the NOS as an explicit component of instruction.

In a study of preservice science methods students, Akerson, Abd-El-Khalick, and Lederman (2000) found that an explicit reflective approach to the NOS was effective in enhancing preservice teachers' ideas about the NOS. Two groups of 25 preservice teachers, enrolled in two sections of the same science methods course, engaged in activities and readings designed to explicitly address the NOS and wrote weekly reflections on the assigned readings and tasks. The preservice teachers' ideas about the

NOS were assessed pre- and post-instruction using a seven-item open-ended questionnaire (p. 300). Initially the teachers were found to have fragmented and inconsistent conceptions of the NOS, and all teachers had naïve views of one or more of the seven aspects of the NOS targeted by the study (p. 305). On the post-instruction questionnaire, the teachers were found to have substantially improved understanding of the target aspects of the NOS (p. 312).

A more recent study of explicit-reflective instruction in the NOS found results consistent with the 2000 study by Akerson, Abd-El-Khalick, and Lederman. A group of 19 preservice teachers enrolled in a science methods class received explicit reflective instruction on the NOS, and their views of the NOS were assessed pre- and post-instruction using an open-ended survey and follow-up interviews (Akerson, Morrison, & McDuffie, 2006). As in the 2000 study, the teachers' conceptions of the NOS were found to improve substantially from pretest to posttest (Akerson, Morrison, & McDuffie, 2006). The researchers in this study, however, assessed the teachers' understanding of the NOS a third time, five months after the conclusion of the science methods class (at the end of the following semester of full-time education coursework). Analysis of this retention test showed that many of the teachers had reverted to some of their original (naïve) ways of thinking about the NOS. Teachers who demonstrated more advanced (less naïve) views of the NOS on the posttest were more likely to retain improvements in their NOS conceptions on the retention test; those whose views were more naïve were more likely to lose some of their gains in NOS understanding. The authors concluded that one course is not enough; in order for teachers to gain and retain a sophisticated understanding of the NOS, repeated exposure to, practice with, and reflection on the NOS is necessary.

Research conducted through the 1970s relied primarily on the use of multiple-choice, ranking, or short-answer surveys to assess students' and teachers' ideas (Lederman, 1992; see also Billeh & Hasan, 1975; Carey & Stauss, 1968, 1970; Kimball, 1967; Mackay, 1971; Miller, 1963), and much of the research since that time has also relied on paper-and-pencil assessments of NOS conceptions (e.g., Akerson, Abd-El-Khalick, & Lederman, 2000; Lawson et al., 2002; Lederman and Zeidler, 1987). A 1990 study of 69 high school students' understanding of the tentative aspect of the NOS cast some doubt on the validity of standard interpretations of such pencil-and-paper assessments. In this study, researchers used a seven-item open-ended questionnaire to assess students' NOS understanding at the beginning and end of an academic year, and then selected 20 of the students for follow-up interviews (Lederman & O'Malley, 1990).

Analysis of students' questionnaire responses indicated that many of the students had absolutist views of science (Lederman & O'Malley, 1990). The follow-up interviews confirmed that students had correctly interpreted the intent of the items on the questionnaire; however, they also showed that the researchers had *not* correctly interpreted many of the students' responses to the items. Specifically, many students had used the term "prove" or "proven" in their responses, which the researchers interpreted as an absolutist stance. In interviews, however, all students were found to use the term "proven" as a synonym for being supported by evidence, without intending to imply absolute truth. Despite having selected students for interviews who were representative of all viewpoints expressed on the questionnaire, the researchers found that all of the students interviewed in fact possessed tentative views of the NOS (p. 235). The researchers concluded that the use of the follow-up interview is essential for interpretation of paper-and-pencil data. Since most of the studies up to that time had not included interviews, the implications of this study for the validity of several decades of research on NOS beliefs are, as the researchers say, “disconcerting at best” (p. 235).

An additional complication to the study of the impact of coursework on student NOS understanding was described by Lising and Elby (2005). While many studies have attempted to characterize the impact of coursework or teaching practices on student understanding of the NOS, the study by Lising and Elby suggests that a student’s understanding of the NOS influences what the student learns from a course. In this case-study of a college student enrolled in introductory physics, the researchers found that the student resisted searching for coherence between formal, mathematical descriptions of phenomena learned in class and informal, intuitive ideas about the same situations. This tendency to avoid seeking coherence in scientific knowledge limited the student’s ability to construct her own understanding of many of the concepts addressed in her physics class. Thus, not only is sophisticated student understanding of the NOS a goal of instruction, but also the actual state of student understanding of the NOS influences the effects of any instruction experienced by the student.

Research on Relationships between Teacher Epistemology and Teaching Practice

In his review of the literature, Lederman (1992) identified two assumptions underlying early research on the NOS. The first, as mentioned above, was that a teacher’s conceptions of the NOS affect the conceptions of his or her students. The second, related, assumption was that the classroom environment established by a teacher, and the actions of the teacher within the classroom, are influenced by his or her conceptions of the NOS (Lederman, 1992, pp. 345-346). Lederman noted that neither of these assumptions was tested in the early research. The first assumption remained essentially untested at the time of his review of the literature (1992), but relates closely to the second assumption. The second assumption was the subject of research in the 1980s and 1990s, with inconclusive results: some studies supported a direct influence of teachers' understanding of the NOS on their classroom practice, and other studies supported the position that there is no influence.

In a 1987 study, for example, Lederman and Zeidler compared experienced teachers' NOS conceptions, as measured by the *Nature of Scientific Knowledge Scale* (NSKS), to 44 classroom variables derived from qualitative analysis of extensive observational field notes taken in the teachers' classrooms. Their sample consisted of 18 10th grade biology teachers with a mean of 15.8 years of experience. The researchers identified four teachers with the most sophisticated understanding of the NOS, as measured by a high score on the NSKS, and four low-scoring teachers with the least sophisticated NOS conceptions. The high-scoring and low-scoring teachers were then compared using the 44 classroom variables. Only one of the 44 classroom variables was found to significantly differentiate between teachers with high and low NSKS scores, and that variable, "Down time" (time spent by students waiting for directions from the teacher), was considered unrelated to NOS understanding. The researchers concluded that their data did not support the idea that teachers' classroom behaviors are directly influenced by their ideas about the NOS.

In a 1991 article, Gallagher analyzed data from a two-year ethnographic study of 27 secondary science teachers to examine questions of how teachers understand the NOS and how this influences their classroom practice. The teachers had a minimum of 10 years experience and most had a major or minor in the subject they were teaching. The researchers found that 25 of the 27 teachers had inadequate conceptions of the NOS. This was seen by the author to be largely the result of college science coursework, which typically focuses on efficient coverage of a large body of scientific knowledge without discussion of how that knowledge was constructed and how it came to be accepted by the scientific community. This point of view is reinforced by the structure of high school science texts, which typically devote hundreds of pages to the current body of scientific knowledge and at most a few chapters to other aspects of science. Gallagher suggested revisions to teacher preparation to address the deficiency in teacher NOS understanding. However, while the classroom actions of the 25 teachers are described, no analysis is included of the teaching behaviors of the two teachers found to have superior understanding of the NOS, and no comparisons of teaching behavior were made between these two teachers and the other 25. Thus, it appears that though the data supported the conclusion that inadequate NOS conceptions lead to teaching inconsistent with the NOS, this study did not provide evidence that a stronger understanding of the NOS by teachers leads to different teaching behaviors.

On the other hand, a qualitative study by Brickhouse (1989) of three teachers with markedly different views of the NOS found a clear relationship between the teachers' views of science and their classroom practice. For example, one teacher viewed scientific theories as truth and the scientific process as a rigid procedure to be repeated many times; his laboratory instruction was highly procedural and focused on following prescribed steps to arrive at the correct result. His teaching behavior was very different from that of the teacher who viewed scientific theories as tools for explaining and predicting observations; her laboratory instruction focused on the use of theory to make predictions and interpret results. In general in this study, a teacher's sophisticated understanding of the NOS was found to correlate with teaching practices consistent with the NOS.

Several more recent studies suggested that a teacher with a sophisticated understanding of the NOS may nevertheless teach science in a way more consistent with naïve conceptions of the NOS. A case study of a high school chemistry class (Tobin & McRobbie, 1997) supported the idea that influences other than the NOS may dominate a teacher's decisions about instruction. In their observations and interviews with the chemistry teacher, the researchers found that the teacher understood science to be inherently tentative and subject to change (p. 361). His classroom teaching, however, consisted almost entirely of memorization of facts and algorithms, more consistent with the idea of science as a body of static and certain facts to be transmitted (p. 365). The researchers found that influences other than ideas about the NOS dominated teaching decisions: the teacher viewed the role of his class primarily as preparing students to pass state chemistry examinations, and to take science courses at the college level, therefore he had a responsibility to transmit a particular set of facts and skills; he saw the curriculum as being determined by others, not within his power to change; and he saw himself as a "guardian of the discipline" of chemistry, with a responsibility to maintaining rigorous instruction (pp. 368-369). All of these factors influenced him to practice teaching methods inconsistent with his own understanding of the NOS.

A study of the NOS understanding and teaching practices of four preservice middle school science teachers in Spain (Mellado, 1997) also showed a lack of correlation between NOS conceptions and teaching methods: student-teachers with more sophisticated understandings of the NOS did not teach in ways consistent with that understanding (pp. 346-347). In another study, Schwartz and Lederman (2002) followed two secondary science teachers from their preservice year, in a MAT program emphasizing explicit instruction of the NOS, through their first year of full-time teaching. This study focused on the teachers' developing understanding of the NOS, as well as the degree to which they were able to include the NOS explicitly in their science instruction. The results of the study suggested that an adequate or sophisticated understanding of the NOS is a necessary, but not sufficient, condition for teachers' implementation of teaching methods consistent with the NOS (Schwartz and Lederman, 2002, p. 230). Deep content knowledge was also necessary; in cases when they had limited knowledge of the science content they were teaching, the teachers were likely to rely on didactic, fact-based teaching methods, even when the lesson addressed ideas about the NOS (Schwartz and Lederman, 2002, p. 229).

Results of a study by Yerrick, Parke, and Nugent (1997) suggested that while a sophisticated understanding of the NOS does not guarantee teaching methods consistent with the NOS, a naïve understanding of the NOS appears to preclude such teaching methods. The context for the study was a two-week summer institute designed to help middle-school science teachers implement new hands-on, inquiry-based science instructional methods consisted with the science education reform movement. Yerrick et al. administered written surveys at the beginning and end of instruction and conducted extensive interviews exploring participants' understanding of the nature of science and science pedagogy. In their analysis, they found that the transformational, constructivist view of science and science teaching which was the framework and primary content of the workshop was interpreted by participants in ways that were incompatible with the intent of the institute.

The researchers found that while teachers participating in the institute adopted language consistent with an understanding of the tentative and constructed nature of science, when questioned further about their use of the terms the teachers were found to have maintained their understanding of science as a body of facts to be transmitted (Yerrick, Parke, & Nugent, 1997, p. 154). Teachers appropriated the language of the institute to describe their existing beliefs and methods, and used instructional strategies consistent with inquiry-based teaching (such as dialog between teacher and students) but used them in ways that fit into a transmission model of teaching (e.g., dialog as a way to assess whether students have acquired the correct body of knowledge). The authors concluded that teachers' understanding of the NOS had a "filtering effect" on their understanding of the goals of science education reform, as represented in the summer institute, and that without a change in their understanding of the NOS, these teachers were unlikely to understand or implement reformed teaching methods (p. 156).

Schwartz and Lederman's 2002 study examined not only whether the teachers were using methods consistent with the NOS, but also whether they were successfully teaching the NOS itself as an explicit component of their lessons. With respect to this goal also, understanding of the NOS was a necessary but not sufficient condition. Again the teachers' understanding of science content was also necessary, as well as the explicit intention (as expressed verbally and in written lesson plans) to include the NOS in their teaching. Without a deep knowledge of science content, the teachers did not recognize where and how the NOS could be taught within the context of science content lessons;

without attention during planning to the elements of the NOS appropriate to the content lessons to be taught, the teachers missed opportunities to teach the NOS in the context of specific content (pp. 229-230).

Abd-El-Khalick, Bell, and Lederman (1998) performed a similar study with 14 preservice secondary science teachers enrolled in a fifth-year MAT program, all of whom demonstrated adequate understanding of the NOS (p. 423). Before beginning their student teaching, all 14 teachers stated they thought it important to teach the NOS in their own classrooms (p. 426). At the conclusion of their 12-week student teaching internships, however, it was found that only three of the teachers had included the NOS explicitly in their science lesson plans (p. 427). The teachers cited a number of constraints as reasons for the absence of the NOS from their lesson plans, including lack of resources, preoccupation with classroom management, and prioritizing of other teaching goals (p. 428).

Twelve of the 14 teachers reported having taught the NOS, but supervisors' field notes and videotapes of the teachers' classrooms showed that teaching of the NOS was much more rare than was reported (Abd-El-Khalick, Bell, & Lederman, 1998, p. 427). This discrepancy was due at least in part to a conflation of the NOS with the processes of science; despite the emphasis of the MAT program on explicit teaching of the NOS, most of the instances teachers gave as examples of having taught the NOS were in fact lessons in which students were engaged in inquiry-based science lessons with no explicit NOS component (Abd-El-Khalick, Bell, & Lederman, 1998, p. 427). As previously discussed, research does not support the assumption that the NOS can be learned implicitly through participation in "doing science." As the authors of this study aptly stated, "the assumption that K-12 students will come to understand the NOS simply through the performance of scientific inquiry and/or investigations is no more valid than assuming that students will learn the details of respiration by breathing" (Abd-El-Khalick, Bell, & Lederman, 1998, p. 430).

While the results of these studies do not suggest a clear relationship between teacher NOS understanding and classroom practice, some trends emerge. Teachers with naïve conceptions of the NOS do not teach in ways that are consistent with the NOS. Teacher with sophisticated understanding of the NOS may or may not teach in ways appropriate to the NOS. At the minimum, these studies support the view that many factors other than teachers' conceptions of the NOS are influential for classroom practice, including content knowledge, teaching context, and curriculum priorities and constraints.

Research on Relationships between Teacher Content Understanding and Teaching Practice

As discussed above, a number of studies exploring the relationship between teacher epistemology and teaching practice found that teacher content knowledge was an important factor influencing teachers' classroom practice. Other researchers setting out to study the relationship between teacher content knowledge and teaching practice have found consistent results.

Cohen (1990) reported a case study of a second-grade teacher who had adopted various new methods and topics in her mathematics instruction, as recommended by the new reform-based texts required by the state of California at the time. The teacher considered her teaching to be completely consistent with the reform effort; Cohen, however, found her teaching to be a complex mixture of reform and traditional instruction. For example, the teacher had adopted “cooperative group learning” by arranging her class in groups of four during math instruction; however, the groups were used only for logistical purposes: no conversation was permitted between group members, and instruction remained entirely teacher-centered (p. 321).

Cohen concluded that the teacher’s understanding of mathematics limited her in two ways. First, her limited background in mathematics prevented her from assisting her students in deepening their own understanding:

> Lacking deep knowledge, Mrs. O was simply unaware of much mathematical content and many ramifications of the material she taught. Many paths to understanding were not taken in her lessons … In these ways and many others, her relatively superficial knowledge of this subject insulated her from even a glimpse of many things she might have done to deepen students’ understanding. (p. 322)

Second, her view of mathematics as a fixed body of knowledge acted as a filter in her adoption of instructional strategies, keeping her class focused on correct answers and preventing her from encouraging discussion and exploration of mathematical ideas.

Cohen’s study demonstrated that a teacher’s limited content knowledge can strongly influence her choice of instructional strategies: despite the interactive nature of the instructional methods advocated by the reform-based texts, Mrs. O’s understanding of mathematics (and mathematics-related pedagogy) led her to implement the methods in a way that removed their interactive and constructivist elements. This suggests not only a lack of content knowledge but also of pedagogical knowledge; Bruner’s interactional and constructivism tenets of education are that student involvement and participation is required for meaningful learning, regardless of the content being studied (Bruner, 1996, pp. 13-19). Lack of content knowledge, however, can lead teachers to choose pedagogically unsound instructional methods even if they demonstrate good pedagogical knowledge when teaching more familiar topics. Shulman (1987) described a case study of a new teacher who taught literature in a style that was "student-centered, discussion-based, … [and] highly interactive" (p. 18). When teaching grammar, a content area with which she was much less comfortable, she replaced her interactive style with "a highly didactic, teacher-directed, swiftly paced combination of lecture and tightly-controlled recitation" (p. 18), and even avoided eye contact with particularly interactive students in order to avoid questions she felt unprepared to answer. Thus even when a teacher understands the pedagogical implications, she may choose poorer instructional methods when faced with teaching content in which she feels under-prepared.

While a lack of content knowledge influences teachers to choose instructional methods inconsistent with constructivism, greater content knowledge does not necessarily lead teachers to choose better instructional strategies. Content knowledge comes in a variety of forms, not all of which are equally useful in preparing a teacher to teach related content. If we assume that the content most important for teachers to understand deeply is that which they will be teaching, then learning more advanced content is only useful to teachers if it can be related to the more introductory content they will teach, either through allowing a deeper understanding of the more introductory content or through making connections between the introductory content and the more advance material. These connections elucidate the conceptual structure and key ideas of the content area, what Shulman described as "the structures of subject matter, the principles of conceptual organization, and the principles of inquiry" (1987, p. 9). As discussed above, advanced coursework does not necessarily lead to deeper understanding of these structures and conceptual principles.

Research on Relationships between Teacher Practice and Student Learning

A large body of research exists on relationships between teacher practice and student learning in science. While some consensus exists that teacher actions in the classroom have a strong influence on student learning (e.g., Greenwald, Hedges, & Laine, 1996; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Wayne & Youngs, 2003; Wright, Horn, & Sanders, 1997), details of what kind of teacher practice is most effective are more difficult to determine. This is due in part to the difficulty of controlling variables, the lack of well-defined categories of teacher practice, and also the impracticality of doing experimental or quasi-experimental studies in educational settings. In general, evidence supports the claim that the kinds of "reformed," constructivist teaching strategies advocated by the *National Science Education Standards* (NRC, 1996) and Project 2061 of the AAAS are more effective than more passive "traditional" forms of science instruction.

A meta-analysis published in 2007 synthesized the results of 61 experimental and quasi-experimental studies on the effects of teaching strategies on student achievement (Schroeder, Scott, Tolson, Huang, & Lee). Researchers grouped the studies, all of which were conducted in the U.S. between 1980 and 2004, into eight major categories based on the dominant alternative teaching strategy examined in the study: questioning strategies; manipulation strategies; enhanced materials strategies; assessment strategies; inquiry strategies; enhanced context strategies; instructional technology strategies; and collaborative learning strategies. The teaching strategies used for the study control groups were not described systematically; however, the authors stated that these strategies were often teacher-centered, rather than the more student-centered alternative strategies used for the experimental groups. The authors noted that the study categories often overlap; also that direct instruction was not included as a category but was the teaching strategy used for the control group in many of the studies. The results of the meta-analysis showed statistically significant effect sizes for each of the teaching strategies examined, suggesting that all of the strategies have positive effects on student learning.

An earlier meta-analysis using the same teaching strategy categories also found positive effect sizes for all alternative strategies (Wise, 1996). The study author described the teaching strategy categories as being unified by the theme of inquiry: only studies in which the researchers identified the teaching strategy as “inquiry” or “discovery” were categorized as such, but all eight categories are characterized by inquiry-oriented, constructivist teaching.

One of the studies categorized under “inquiry strategies” in the 2007 meta-analysis (Schroeder et al.) reported on the results of a three-year middle school science systemic reform effort carried out in Detroit between 1998 and 2001 (Marx et al., 2004). As part of the reform effort, a group of middle school science teachers was recruited to implement newly developed curriculum units that were developed by the researchers in collaboration with Detroit Public Schools and were designed to be centered on context-rich driving questions; inquiry-based and technology-infused; and strongly aligned with the *Benchmarks for Science Literacy* (AAAS, 1993), *National Science Education Standards* (NRC, 1996), and the Detroit curriculum framework. Teachers participated in extensive professional development over the three years of the project, and were involved in the refinement of both the curriculum and the assessments used to measure student learning. Four curriculum units were used; participation numbers for each year and unit ranged from 110 students taught by 3 teachers to 1203 students taught by 14 teachers (Marx et al., p. 1069).

The researchers analyzed student pretest and posttest scores on the collaboratively developed assessments to determine the effectiveness of the enacted curriculum for student learning of both content and process skills (Marx et al., 2004). Students were found to have achieved statistically significant gains on all units in all three years; also, the effect size of student gains increased each year, indicating increasing benefits for students with sustained attention to the curriculum and related professional development. The researchers concluded that inquiry-based science instruction consistent with national science education reform efforts can succeed in urban school districts with high levels of poverty and teacher turnover. It should be noted that this instructional method was not compared to other methods of teaching the same material; rather, the comparison was of student knowledge and skills before and after each curriculum unit. This study therefore demonstrated the success of the implemented curriculum, but did not compare the success of this strategy to that of other instructional methods. The authors also noted that their study could not differentiate the relative effects of the many components of the curriculum implementation, which included systemic support for reform, professional development for teachers, collaborative curriculum design and modification, and the use of technology. This supports the statement made by Schroeder et al. (2007) that the categories of teaching strategy used in their meta-analysis are not independent, and in fact many studies involve strategies fitting into more than one category.

A smaller-scale study categorized under “questioning strategies” by Schroeder et al. (2007) used a two-day 10^{th} grade biology lesson for 107 students to examine the effects of orienting questions on student learning (Osman & Hannafin, 1994). Students were stratified into low and high levels of prior knowledge, as measured by the science score on the California Test of Basic Skills, and then randomly assigned to treatment groups. The lesson consisted of a reading assignment in the area of genetics, with or without embedded questions designed to activate relevant conceptual prior knowledge; the posttest contained both factual questions and problem-solving questions. The researchers found that both high- and low-prior knowledge students who received orienting questions had significantly higher posttest scores than those students who received reading assignments without orienting questions.

High school biology students were also the participants in a study categorized under “manipulation strategies” in the Schroeder et al. (2007) meta-analysis (Matthews & McLaughlin, 1994). In this study, 46 senior biology students in two classes participated in a four-month unit on cell biology; one class (the experimental group) completed weekly hands-on student-centered laboratory activities, and the other class (the comparison group) did not. The researchers found that students who participated in laboratory activities scored significantly higher on the unit posttest than did students in the comparison group. Unfortunately, the study was not described in sufficient detail to be replicable: instructional methods were not described beyond the inclusion of hands-on laboratory activities for one of the groups; randomization level is not described; and possible confounding variables for validity and reliability are not discussed.

Another approach to the study of teaching methodology and student learning was used by the Evaluation Facilitation Group of the ACEPT group at Arizona State University (Sawada et al., 2002). Rather than group teachers into categories by teaching strategy, the Evaluation Facilitation Group developed an observation protocol to make a holistic measurement of the degree to which observed teaching of mathematics and science is "reformed." As described above, the group defined reformed teaching to be characterized by the constructivist, student-centered methods advocated by the NRC (e.g., 1996), AAAS Project 2061 (e.g., 1989, 1993), and the National Council of Teachers of Mathematics. The resulting instrument, the RTOP, had very high interrater reliability for trained observers, based on 24 paired observations in college physics, physical science, and mathematics courses. The group then used the RTOP to rate 16 university professors whose students took standardized content pretests and posttests; teacher RTOP score was found to correlate strongly with student learning gains on standardized content assessments. For example, in the six physical science classes studied, the correlation coefficient between teacher RTOP and average student normalized learning gain was 0.88 (Sawada et al., p. 249).

The Arizona Evaluation Facilitation Group also studied a large-enrolment introductory biology course with a focus on scientific reasoning skills (Lawson et al., 2002) and found similar results. Students in the biology class attended the same lecture sections and were divided into weekly two-hour laboratory sections taught by nine different graduate teaching assistants (TAs). The TAs were introduced to reformed teaching methods and were provided with inquiry-based labs and ongoing two-hour weekly meetings during the semester; nonetheless, mean TA RTOP scores for the semester ranged from 42 to 90 out of 100 possible points (Lawson et al., p. 391). A test of scientific reasoning skills was used as a pretest and posttest for students. The researchers found a strong correlation between TA RTOP score and student learning gains on the test of scientific reasoning. While the students studied in this research were at the college level, all courses were at the introductory level and therefore similar in content to what might be seen in high school science and mathematics courses.

Another study by the ACEPT group found similar results for secondary science students. As described above, the 2003 study by Adamson et al. characterized secondary teachers by the number of ACEPT-influenced courses they experienced as undergraduates, and found that ACEPT-influenced teachers had higher RTOP scores than non-ACEPT teachers. Furthermore, the students of ACEPT-influenced biology teachers were found to have higher biology achievement than the students of non-ACEPT teachers, supporting the idea that the teacher behavior described by the RTOP results in more student learning. This was true for three of the four subscales of the achievement measure (NOS, scientific reasoning, and biology concepts) as well as for overall score, though differences on the subscale scores were statistically significant for some but not all student groups. The researchers also noted their multiple regression analysis showed that student scores were better predicted by the level of ACEPT influence of the teacher (zero, one, or two or more courses) than by teacher RTOP score (p. 950), suggesting that even the detailed observational evidence gathered with the RTOP misses some relevant aspect of teacher practice that is influenced by experiencing reform teaching, and in turn influences student learning.

Summary of Salient Findings

The relationship between teacher characteristics and student learning is complex and not well understood. Little empirical research has been done that directly relates teacher characteristics to student learning (Greenwald et al., 1996; Shulman, 2002; Wilson et al., 2002). Those studies that have been done have found complex mixed or inconclusive results (Greenwald et al., 1996; Monk, 1994; Wayne & Youngs, 2003; Wilson et al., 2002), not easily interpreted; for example, Monk (1994) found a mix of positive and negative correlations between teacher coursework in science and mathematics and learning gains by students.

Research has been done examining intermediate steps between teacher characteristics and student learning. For example, researchers in physics education have found that instructional method is related to student learning: research-based, inquiry-oriented instruction leads to greater improvement in conceptual understanding than do lecture-based instruction and traditional problem-solving activities (Hake, 1998; Kim & Pak, 2002; Mazur, 1997; McDermott & Shaffer, 1992; McDermott, Shaffer, & Constantinou, 2000). In addition, instructional method may negatively influence student beliefs about the nature of science and of teaching and learning about science (AAAS, 1998; Grossman & Stodolsky, 1995; McDermott et. al, 2000). Deep content knowledge on the part of teachers is a necessary, but not sufficient, condition for the kind of teaching advocated by the current reform movement (Cohen, 1990; Shulman, 1987).

Research on students' and teachers' conceptions about the nature of science has found similarly mixed results. Research in the 1960s and 70s suggested that on average, both teachers and students had poor understanding of the nature of science (Lederman, 1992). Despite efforts in the post-Sputnik years to develop pre-college science curricula consistent with the nature of science, students who participated in these curricula did not develop the desired understanding of the nature of science (Abd-El-Khalick, Bell, & Lederman, 1998). Teachers' conceptions of the nature of science did not correlate with their college science credits, college grade-point average, or other relevant academic variables (Lederman, 1992). Some studies have provided evidence that how teachers think about science influences their behavior in the classroom. Specifically, if teachers understand science or mathematics as a body of facts to be transmitted, they are unlikely to adopt reform-oriented pedagogical strategies even if trained to do so (Cohen, 1990; Yerrick et al., 1997). More recent research has focused on improving preservice and in-service teachers' conceptions, with some success (Akerson, Abd-El-Khalick, & Lederman, 2000; Akerson, Morrison, & McDuffie, 2006). While a sophisticated understanding of the nature of science was not found to guarantee teaching methods consistent with the nature of science (Lederman, 1992; Mellado, 1997; Schwartz & Lederman, 2002; Tobin & McRobbie, 1997), a naïve understanding appears to preclude such teaching methods (Cohen, 1990; Grossman & Stodalsky, 1995; Yerrick et al., 1997).

The goal of science literacy for all requires examination of student learning. The goal of improved K-12 science education requires research on how teacher characteristics are related to student learning, so that appropriate education and support can be provided for both preservice and in-service science teachers. This study is intended to add to the limited body of existing research on relationships between teacher characteristics and student learning. In addition to teacher characteristics included in previous studies, such as teacher coursework and years of experience, this study also examines teacher epistemological beliefs, including teacher beliefs about the nature of science.

The specific research method, measures, and procedures used in this study are described in chapter 3.

Chapter 3: Research Methods

Research Questions

This study was designed to contribute to the research base relating teacher characteristics to student learning. Specifically, the study attempted to address the following research questions:

1. Is there a relationship between student gain on the researcher-designed content tests and student scores on the science portion of the Washington State Assessment of Student Learning (WASL)?
2. Beyond any relationship between students' gain on the researcher-designed content tests and their scores on the Science WASL, is there a relationship between students' gain on the researcher-designed content tests and teachers' epistemological beliefs, coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, coursework in science content or science education, or years of experience?
3. Is there a relationship between teachers' epistemological beliefs and their years of teaching experience?
4. Beyond any relationship between teachers' epistemological beliefs and their years of teaching experience, is there a relationship between teachers' epistemological beliefs and their coursework in science content or science education, in inquiry-oriented teaching strategies, or in formative assessment and evaluation of students' prior knowledge?

Research Design

This investigation into relationships between teacher characteristics and student learning is a correlational study using a descriptive design. The investigation is correlational in that it describes relationships between the variables measured. The investigation is descriptive in that it relates measures of already existing teacher characteristics to measures of student achievement.

Participants

Sampling Methods

This study was conducted as part of a larger five-year research effort undertaken by members of the SPU Physics Department and researchers from FACET Innovations, LLC. The research team has an ongoing relationship with the science curriculum developer in the school district in question, as well as with many of the science teachers in the district. Members of the research team have provided a number of professional development workshops and courses for science teachers in the district, both in subject areas related to the research effort and in other areas as requested by the science curriculum developer. In this study, our focus has been on 8th grade science teachers in the district and their students.

For the 2005-2006 academic year, the first year of the larger research effort and the year during which the student data described in this study was collected, all 8th grade science classes in the district used the curriculum *Properties of Matter*, from the National Science Resources Center project “Science and Technology Concepts for Middle Schools” (National Academy of Sciences, 2000). This curriculum unit focuses on measurement and interpretation of properties of matter, such as mass, volume, and density. This was the first of two curriculum units taught during 8^{th} grade, and was used for approximately the first half of the academic year. Members of the research team conducted professional development workshops for school district teachers focused on content relevant to the *Properties of Matter* curriculum, including a workshop in which results of the researcher-developed pretest and posttest were presented and discussed in the context of instructional decisions and curriculum adaptation and revision.

The school district has a system of common exams for all curriculum units, including pretests in some cases. For the *Properties of Matter* curriculum, all 8^{th} grade teachers in the district administered the researcher-developed pretest and posttest. In addition, all students in the district participated in the WASL in the third week of April 2006, which for 8^{th} grade students includes a science assessment.

Student data was collected for the students of 13 teachers during the 2005-2006 academic year. Teacher data was collected the following academic year. Several teachers left the school district before data collection was complete and were difficult to locate during teacher data collection. Efforts to locate those teachers and solicit their participation continued through summer of 2008. Complete data was obtained for 9 of the 13 teachers. One of the remaining teachers declined to participate, and three could not be located.

Description of Participants

This study was conducted in a large suburban school district in the Pacific Northwest. The total population of the district is approximately 115,000, and the total student enrollment in the district approximately 16,000 students. The school district contains 16 regular elementary schools, five regular middle schools, and four comprehensive high schools, as well as several alternative schools including a K-12 home school cooperative, a 6-12 international school, and a middle school and high school designed to serve students who have not experienced success in traditional educational settings. Participants in this study included students from all five regular middle schools, as well as a few students from the alternative middle school. The district reported the following student ethnographic data for the 2005-2006 academic year: 57% White, 24% Asian, 8% Hispanic, 8% Multi-Ethnic, and 3% African American. According to state data, 19% of students in the district qualified for free or reduced price lunch in October 2005.

Initially teachers were contacted from a list of all middle school science teachers provided by the science curriculum developer. Follow-up contact was focused on 8th grade teachers whose students participated in the pretest and posttest. A total of 22 teachers completed the Teacher Questionnaire. Of these, 15 also completed the Epistemological Beliefs Assessment for Physical Science (EBAPS). Of these 15 teachers, nine were 8th grade teachers whose students participated in the researcher-developed pretest and posttest.

Procedures

The pretest and posttest on properties of matter content were written by members of the research team in collaboration with the school district science curriculum developer and administered by teachers as part of regular instruction. Both tests were administered by computer, typically in a computer laboratory classroom separate from the normal science classroom. According to the science curriculum developer, relocating a class to the computer lab is unusual but not unprecedented, taking place a few times each year; students would also have experienced the computer lab in previous years, so the setting would not have been unfamiliar to most students (personal communication, September 4, 2008). Computer use in the classroom is not unusual; all schools had computer carts with 16 laptop computers designated for science classroom use on a regular basis. Results of the tests were analyzed by the research team and presented to the teachers, followed by discussion of implications of the results and possible instructional modifications.

The Teacher Questionnaire was administered to teachers using the online survey website SurveyMonkey (http://www.surveymonkey.com). Teachers were initially contacted by email in December 2006 with the request to fill out the online questionnaire. Teachers were asked to sign a consent form and were informed that their responses would be confidential and all identifying information would be removed from any published results. Follow-up emails and phone calls continued through the following autumn. Teachers who completed the questionnaire were sent a $25 gift card.

The EBAPS was administered on paper to teachers participating in a professional development workshop at Seattle Pacific University in March 2007. Teachers who did not participate in the March 2007 workshop were contacted by email or phone over the following months to request their participation. Those who agreed to participate in the study completed the EBAPS via SurveyMonkey. Follow-up emails and phone calls continued through autumn of 2007.

The WASL is the state standardized assessment used to meet state and federal testing requirements. All students in public schools in grades 3 through 10 in Washington State take WASL exams each spring. Tests in reading and mathematics are taken every year; writing is taken in grades 4, 7, 9 and 10; science is taken in grades 5, 8, and 10 (*WASL Timelines and Calendars,* n.d.). Students in 8th grade therefore take WASL tests in reading, mathematics, and science. In this district, science was taken last, after reading and mathematics. The WASL is administered by school personnel and has no time limit (*Washington Assessment of Student Learning,* n.d.).

To access student WASL results, permission was obtained from the school district and appropriate documentation was filed with the school district and with the Washington Office of Superintendent of Public Instruction (OSPI). Personnel in the Office of Student Information and Assessment Operations at OSPI compiled school district WASL data, including student ID numbers, so that the student WASL data could be matched with pretest and posttest data and student could be matched with science teachers.

Measures

Teacher Questionnaire

The Teacher Questionnaire is a self-report instrument that asks questions about teachers' background, educational beliefs, teaching practice, and perceived teaching environment (see Appendix A for complete questionnaire). It is based in part on the teacher questionnaire developed by Horizon Research, Inc., for assessment of the National Science Foundation Local Systemic Change program. The section on background was expanded to include information about coursework preparation in specific areas, including physical science, life science, science education, and use of inquiry-oriented teaching strategies. To facilitate administration, the questionnaire was converted to an electronic format using the online survey website SurveyMonkey.

EBAPS

The EBAPS was developed by Andrew Elby and colleagues at the University of California, Berkeley (*Epistemological Beliefs*, n.d.). It is a forced-choice instrument designed to probe beliefs about epistemology in the context of physical science (see Appendix B for the complete instrument). It was originally designed to be used in high school and college introductory courses in physics, physical science, and chemistry. The instrument contains three types of questions: Likert-scale items on which participants agree or disagree with statements; multiple-choice items; and “debate items,” in which participants read a short debate between two characters and indicate their level of agreement with the ideas expressed by each.

The authors described five subscales within the EBAPS instrument: “structure of scientific knowledge,” “nature of knowing and learning,” “real-life applicability,” “evolving knowledge,” and “source of ability to learn.” All questions on the instrument are explicitly contextualized in the physical sciences. For example, questions assigned to the subscale “nature of knowing and learning” probe participants’ beliefs about learning science.

Each EBAPS item is scored on a scale of 0 to 4, with 4 representing the most sophisticated (and therefore most correct) response. The scoring scheme is non-linear and takes into account variations in the level of sophistication of possible responses due to differences in question content and type (*Epistemological Beliefs*, n.d.). For example, on Likert-scale items, the neutral response (“neither agree nor disagree”) is scored from 1 to 2.5 points, depending on the question (*Which EBAPS Items*, n.d.). For a complete list of questions including subscale assignment and scoring weights, see Appendix C. A subscale score is the simple average of scores from all questions on the subscale; a total score is the simple average of all EBAPS items (*Epistemological Beliefs*, n.d.).

Although the EBAPS is a well-known instrument within the community of physics education researchers (see, e.g., Adams et al., 2006), the instrument itself has not been the subject of any peer-reviewed publications. Information about validity and reliability is published in the document *The Idea behind the EBAPS* (n.d.), on the website of Elby. Validity testing was focused on how well items tested students' epistemological beliefs separately from any expectations related to specific science courses or learning goals. Initial revisions were made based on pilot testing and informal feedback. A larger validity study was then completed using data from approximately 100 students from six community colleges in northern California. Students completed the assessment and wrote explanations for their selections; their responses were then analyzed and coded for epistemological and non-epistemological content, and this analysis was used to revise items that triggered non-epistemological reasoning in student responses.

The reliability of a test is a characterization of how accurately and consistently the test assesses the true value of the characteristic it is attempting to measure (Gall, Gall, & Borg, 2003, p. 195). For example, a personality test may attempt to measure whether a person is introverted or extroverted; in this case, the test would contain a set of questions designed to assess the level of introversion or extroversion of the test-taker. The reliability of these test items could be evaluated by comparing a person's responses on half the questions to the same person's responses to the other half. If the quality of introversion or extroversion is a relatively stable personality trait—that is, if a person has a "true" value of degree of introversion or extroversion—then the responses on the two half-sets of questions measuring that personality trait should be consistent with each other. Any inconsistency between responses would be deemed a measurement error in the test itself, rather than reflecting real differences in the personality of the test-taker. Thus, the assumption of trait stability or consistency is central to the measurement of test reliability.

The EBAPS subscales, however, were not designed to assess stable characteristics or beliefs of the test-taker; instead, they can be viewed as targets for understanding (*The Idea behind the EBAPS*, n.d., section 3.3). For example, science course goals may include the development of coherent, consistent student understanding of the structure of scientific knowledge. This does not mean that students are successful in developing such an understanding. Inconsistent student responses to different questions designed to assess understanding of the structure of scientific knowledge, therefore, do not necessarily indicate that the questions are flawed; rather, they show that the test-taker may not have a coherent and consistent understanding of the concepts the questions are designed to assess. Because of this distinction, the authors of the EBAPS chose not to conduct reliability testing on the instrument (*The Idea behind the EBAPS*, n.d., section 3.3).

Science WASL

According to the Washington OSPI website, the WASL is the product of an education reform movement which began in 1993 when the Washington State legislature required the state to create a set of learning standards and a testing system to measure student learning. Essential Academic Learning Requirements (EALRs) for science were adopted in 1998, and the first Science WASL was given in 2003 (Science Assessment Team, 2007, p. I.3). The science content of the WASL is divided into three content strands: Systems of Science, Inquiry in Science, and Applications of Science. The points available on the test are allocated 40% to each of the first two strands and 20% to the Applications strand (Science Assessment Team, 2007, p. IV.5). The science WASL is composed of three types of questions: multiple-choice (45% of available points), short answer (36%), and extended response (19%) (Science Assessment Team, 2007, p. IV.2).

Short-answer and extended-response questions are scored by a team of trained and supervised scorers using rubrics that have undergone an extensive review process (Science Assessment Team, 2007, p. III.13). Total raw scores are converted to scaled scores using a linear transformation in which specified scaled scores are fixed to cut-points for “basic” and “proficient” levels of performance. Final scaled scores are classified according to four categories: “below basic,” “basic,” “proficient,” and “advanced,” where students in the “proficient” and “advanced” categories are considered to have met the performance standard for the test (Pearson Educational Measurement, 2007, p. 40).

As described by the Science Assessment Team of the Washington OSPI (2007), the process of developing the science WASL included multiple reviews by content area experts, test developers, and professional educators in order to ensure content validity of all test items. The technical report on the 2006 8th grade WASL (Pearson Educational Measurement, 2007) reported on internal consistency measures of reliability. On short-answer and extended-response WASL items, inter-rater reliability is established through various procedures including double-scoring, back-scoring by scoring supervisors, and blindly inserted validity papers (Office of Superintendent of Public Instruction, 2006). Inter-rater reliability on these questions was found to range from 80% to 96% for exact score matches and from 96% to 100% for exact and adjacent score matches (Pearson Educational Measurement, 2007, p. 54). The authors of the technical report noted that conditional standard error of measurement of the scale scores was large enough to require caution in interpretation of individual student scores (Pearson Educational Measurement, 2007, p. 38). Overall, however, the results suggest strong validity and reliability for the science WASL.

The Cronbach's alpha coefficient was calculated for each strand of the science WASL and was found to be .79 for the Systems of Science strand, .80 for Inquiry in Science, and .56 for Applications of Science (Pearson Educational Measurement, 2007, p. 36). While the coefficient for the third strand seems somewhat low, the authors noted that the heterogeneity of test items may cause the Cronbach's alpha coefficient to under-estimate the true reliability of the test (Pearson Educational Measurement, 2007, p. 34). The Cronbach's alpha results were interpreted as evidence of acceptable levels of internal consistency.

Researcher-Designed Pretest and Posttest

Both the pretest and the posttest were written and reviewed by members of the research team. Members of the research team represented a wide variety of relevant experience, including Seattle Pacific University faculty in physics and science education; research scientists from FACET Innovations, LLC, an educational research and development company specializing in science education; active and retired secondary science teachers; and a veteran elementary teacher and science specialist. Both sets of test questions were also revised based on extensive review and feedback from the school district science curriculum developer. The posttest questions were written after the pretest had been administered, and were influenced by student responses to pretest questions. Some pretest questions were found to have high student success rates before instruction, and therefore the research team chose not to include parallel questions on the posttest. Questions were added to the posttest beyond the content covered on the pretest in order to extend the range of student ideas explored and increase the difficulty of the test. It is important to note that in addition to assessing student understanding, one purpose of the tests was to explore student ideas in the content realm for the purpose of informing the development of a formative assessment tool.

The pretest contained 12 questions; the posttest contained 17 questions. Both tests consisted of two types of question: multiple-choice, for which students selected the best choice from a list; and multiple-response, for which students selected all applicable choices from a list. In addition, several questions on each test were followed by an open-ended prompt asking students to explain their answer.

The development process for the tests was focused on content validity. Reviews by K-12 teachers on the research team helped to ensure age-appropriate language and question contexts. Additional reviews were completed by two university content experts (one in physics and one in chemistry) and one secondary science teacher unaffiliated with the research team. Based on feedback from these external reviewers, one pair of questions was removed from the analysis due to possible ambiguity in wording of the posttest question. In addition, for one multiple-response question on the pretest, one choice was considered ambiguous: with appropriate reasoning, it could be considered either applicable or inapplicable. This choice was removed from the analysis; the question and other responses remained part of the analysis. Content validity and clarity of the remaining questions was affirmed by the reviewers.

Eight of the posttest questions tested the same concepts as eight questions from the pretest. These eight paired questions were initially selected for the analysis of student learning gains. Due to validity concerns for one posttest question, as discussed above, one pair of questions was eliminated from the analysis; seven paired questions were included in the final analysis of student learning gains (see Appendix D).

For multiple-choice questions, each student response was scored as a 1 (for the correct response) or a 0 (for any incorrect response). For multiple-response questions, each choice was scored separately and was worth a fraction of 1 point. For example, for a multiple-response question in which students were instructed to select all correct choices from a list of five, each choice was scored as 1/5 of 1 point. Student pretest and posttest scores were calculated by summing the student scores for each question. Student learning gains were calculated by subtracting each student's pretest score from the student's posttest score.

Data Analysis

Descriptive and inferential statistics were calculated to address the research questions. The data were analyzed to determine whether the assumptions of parametric procedures were met. Inferential statistics were computed using hierarchical multiple regression. Multiple regression was chosen in order to generate expressions showing the amount of relationship (correlation) between several predictor variables (e.g., student WASL score, teacher coursework in science) and a continuous criterion variable (e.g., student learning gain as measured by the researcher-designed tests). In multiple regression, the criterion variable must be continuous; predictor variables may be continuous, ordinal, or categorical (Tabachnick & Fidell, 2007, p. 119). Hierarchical multiple regression was chosen to allow the predictor variables to be entered in two blocks in specified order, in order to assess the contributions of the second block of predictor variables over and above the first block.

Initial analysis revealed multicollinearity between the three variables related to teacher coursework preparation: coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, and coursework in science content or science education. These three variables were therefore combined into one coursework variable, and the research questions were revised to reflect this change. Descriptive and inferential statistics were re-calculated to address the revised research questions.

A hierarchical multiple regression analysis was performed for each research question using the linear regression program of the SPSS software package (version 15.0). Results of each statistical analysis are presented in chapter 4. A discussion of the practical significance of the results is presented in chapter 5, along with limitations of the current study and directions for future research.

Chapter 4: Results

Introduction and Revised Research Questions

The results of this study are presented by research questions. Descriptive statistics are provided for all variables included in the study. Descriptive statistics given include measures of the central tendency, variability, and normality of the data sets. Assumptions underlying the relevant inferential statistical procedures are reviewed and the data used in this study is discussed in light of these assumptions. Revisions made to the original research questions based on initial data analysis are described. Finally, inferential statistics are presented and discussed.

Research questions 1 and 2 examine student performance gain on the researcher-designed pretest and posttest. Research question 1 asks whether there is a relationship between student performance gain and scores on the science portion of the Washington Assessment of Student Learning (WASL). Research question 2 asks whether, beyond the relationship described in research question 1, there is a relationship between student performance gain and the following teacher characteristics: epistemological beliefs, coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, coursework in science content or science education, and years of experience.

Research questions 3 and 4 examine teacher characteristics. Research question 3 asks whether there is a relationship between teachers' epistemological beliefs and their years of teaching experience. Research question 4 asks whether, beyond the relationship described in research question 3, there is a relationship between teachers' epistemological beliefs and the following characteristics: coursework in science content or science education, coursework in inquiry-oriented teaching strategies, and coursework in formative assessment and evaluation of students' prior knowledge.

Descriptive and inferential statistics were calculated for each pair of research questions and the associated data sets. Each pair of research questions was evaluated using hierarchical multiple regression. Multiple regression produces a prediction equation relating the criterion variable to the set of predictor variables. The hierarchical multiple regression procedure allows the researcher to select the order of input for the predictor variables, thus permitting the researcher to statistically control for the effects of specific predictor variables. The hierarchical multiple regression analyses were performed using the linear regression program of the SPSS software package (version 15.0).

Analysis conducted on the data for the original research questions revealed high levels of collinearity between the three teacher coursework variables: coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, and coursework in science content or science education. The Pearson correlation coefficient between formative assessment coursework and inquiry-oriented coursework was $r = .783$ ($p < .001$); correlations between science and science education coursework and formative assessment coursework or inquiry-oriented coursework were $r = .587$ ($p < .05$) and $r = .501$ ($p < .05$), respectively. Using highly correlated predictor variables in a multiple regression analysis can reduce the probability of finding a statistically significant result (Licht, 1995, p. 45; Tabachnick & Fidell, 2007, p. 84). Tabachnick and Fidell (2007) have suggested that bivariate correlations of $r = .70$ or more between predictor variables in a multiple regression analysis should probably be avoided by excluding one of the two variables or combining them into a composite score (p. 90).

In order to avoid highly correlated predictor variables, the decision was made to combine the three variables related to teacher coursework into one new variable: teacher total combined coursework credits in the areas of science content, science education, inquiry-oriented teaching strategies, and formative assessment and evaluation of students' prior knowledge.

The decision to combine teacher coursework data into one variable necessitated revision of two of the four original research questions. The revised research questions are given below.

Research questions 1 and 2 examine student performance gain on the researcher-designed pretest and posttest. Research question 1 asks whether there is a relationship between student performance gain and scores on the Science WASL. This research question was not revised. Research question 2 was revised. The new research question 2 asks whether, beyond the relationship described in research question 1, there is a relationship between student performance gain and the following teacher characteristics: epistemological beliefs, combined coursework, and years of experience.

Research questions 3 and 4 examine teacher characteristics. Research question 3 asks whether there is a relationship between teachers' epistemological beliefs and their years of teaching experience. This research question was not revised. Research question 4 was revised. The new research question 4 asks whether, beyond the relationship described in research question 3, there is a relationship between teachers' epistemological beliefs and their combined coursework.

Descriptive Statistics

Students

Descriptive statistics for student data are provided in Table 1. Student WASL scores range from 4 to 59 out of a maximum possible score of 62 (Science Assessment Team, 2007, p. IV.2), with a mean of 37.9. Both the skewness and kurtosis statistics fall between plus one and minus one. Student gain score on the researcher-developed tests was calculated by subtracting the pretest score (M = 3.11, SD = 1.06) from the posttest score (M = 4.11, SD = 1.00) for each student. As shown in Table 1, student gain score ranges from -3.14 to 4.99, with a mean of 1.00 and standard deviation of 1.42. Both skewness and kurtosis statistics fall between plus one and minus one.

Table 1

Descriptive Statistics for Student Data

	WASL raw score	Pre-post gain
N	1,250	823
Mean	37.9	1.00
Std. Deviation	11.1	1.42
Minimum	4	-3.14
Maximum	59	4.99
Skewness	-0.72	0.02
Kurtosis	-0.05	-0.31

Because the multiple regression analysis uses only cases for which complete data are available, a subset of 642 of the student cases represented in Table 1 are used in the multiple regression analysis. Though the use of a convenience sample does not allow for choosing a representative subset of cases for analysis, the descriptive statistics for this subset (N = 642) are similar to those seen in Table 1 both for student gain score (M = .90, SD = 1.39) and student WASL raw score (M = 37.2, SD = 10.5); for the subset as well as the larger sample, both the skewness and kurtosis statistics fall between plus one and minus one.

Teachers

Descriptive statistics for participating teachers are described below and in Tables 2 and 3. All data with the exception of teacher score on the Epistemological Beliefs Assessment for Physical Science (EBAPS) are taken from teacher responses on the self-report Teacher Questionnaire (see Appendix A). Teacher Questionnaire data was received from 22 teachers; of these, 15 also completed the EBAPS. Descriptive statistics are reported for the 15 teachers with complete data.

Teachers were asked to select from five possible responses to describe their years of teaching experience: More than 20 years; 10 to 20 years; 5 to 9 years; 2 to 4 years; or 1 year or less. Teacher responses are summarized in Table 2. As can be seen from Table 2, approximately half of the participating teachers had between 5 and 9 years of experience.

Table 2

Teacher Years of Experience

Experience	*N*
1 year or less	1
2 to 4 years	1
5 to 9 years	7
10 to 20 years	5
More than 20 years	1
Total	15

Two sets of questions on the Teacher Questionnaire asked teachers to report numbers of coursework credit-hours in various categories. The first set of questions asked the responding teacher to report credit-hours taken before the teacher began full-time teaching, including both undergraduate and graduate courses; the second set of questions asked the responding teacher to report credit-hours taken after the teacher began full-time teaching, including both undergraduate and graduate courses. In both sets of questions, respondents were asked to report credit-hours of coursework in the following categories: mathematics; physical science; life science; mathematics education; science education; coursework focused on the use of questioning, discussion, and inquiry-oriented teaching strategies; coursework focused on using technology in the classroom; and coursework focused on the use of formative assessment and evaluation of students' prior knowledge to inform instruction. For each category, teachers selected from six possible responses for number of credit-hours: 0; 1-3; 4-9; 10-15; 16-30; or more than 30.

Because this study does not seek to differentiate between coursework taken before and after teachers began full-time practice, it was necessary to combine the two sets of responses. In addition, for this study several categories of coursework were combined into a more broadly-defined single variable. Initially, the categories of physical science, life science, and science education were combined into one coursework category, described as coursework in science content or science education. Initial analysis of the data revealed unacceptably high bivariate correlation coefficients between this coursework variable and the two additional coursework variables, as described above. Therefore, teacher data on coursework in science content and science education were further combined with teacher data on two additional coursework variables: coursework focused on the use of questioning, discussion, and inquiry-oriented teaching strategies; and coursework focused on using technology in the classroom; and coursework focused on the use of formative assessment and evaluation of students' prior knowledge to inform instruction.

In order to combine teacher responses to multiple coursework-related questions, each teacher response was translated into the average of the high and low ends of the credit-hour range selected. For example, a selection of the response "1-3" would translate to 2 credit-hours. This allowed each teacher response to be represented as a single number. Teacher responses to multiple questions were then combined by finding the sum of the numbers representing each response. After all selected responses were combined into a single number for each teacher, the resulting data were categorized into ranges of total coursework credits. The results are summarized in Table 3. As can be seen from Table 2, teachers were approximately normally distributed across the categories.

Table 3

Teacher Coursework in Science, Science Education, Inquiry, and Formative Assessment

Coursework credits	*N*
60 or fewer	1
61 to 80	4
81 to 120	6
121 to 170	3
More than 170	1
Total	15

The EBAPS was administered online through a survey tool and scored using the scoring template provided by the authors (see Appendix C). Total possible overall score ranges from 0 (unsophisticated) to 4 (highly sophisticated). Teacher scores range from 2.95 to 3.58, with a mean of 3.25 and standard deviation of 0.19. Both the skewness statistic (0.39) and kurtosis statistic (-0.53) fall within the range of plus or minus one. These scores indicate that on average, the teachers participating in the study had fairly sophisticated epistemological beliefs.

Assumptions of Inferential Procedures

Before reporting the model resulting from the multiple regression analysis, it is necessary to examine the results to assess whether the model fits the data well and whether the underlying assumptions of the analytic procedure have been met. Some assumptions concern the raw data; others concern the residuals. Residuals are the differences between the observed value of the criterion variable and the value predicted by the regression model for each case (Field, 2005, p. 163).

Absence of Multicolliniarity

Multiple regression analysis produces useful results only if the predictors are correlated with the criterion variable and not too highly correlated with each other. As correlations between predictor variables increase in size, the regression equation produced by the multiple regression analysis becomes less descriptive of relationships between predictors and the criterion. In the absence of correlation between predictor variables, the relative sizes of the standardized coefficients of each predictor in the regression equation can be interpreted as indicated the degree of importance each variable has in predicting the criterion (Licht, 1995, p. 47). However, when predictor variables have larger correlations with each other, it is more likely that some of the variance in the criterion is shared by the correlated predictor variables, rather than being attributable to a single predictor. In this case, the multiple regression analysis does not attribute the shared variance to either predictor variable; therefore, the regression coefficients of the predictor variables are affected by the relationship between predictors, and can no longer be said to represent only the relationship between each predictor and the criterion (Licht, 1995, pp. 46-47). In addition, strong correlation between predictor variables increases variance in the regression coefficients produced by the multiple regression analysis, which increases the likelihood of sample-specific results (Field, 2005, p. 175). According to Licht (1995), the ideal predictor variables in a multiple regression analysis are be highly correlated with the criterion variable and uncorrelated with the other predictor variables (p. 46; see also Tabachnick & Fidell, 2007, p. 122).

The correlation matrix for the multiple regression analysis of student variables is shown in Table 4. The correlation matrix shows acceptably low correlation between each pair of predictor variables. The highest correlation is between teacher years of experience and teacher combined coursework credits ($r = .58$, $p < .001$); this is not unexpected, since in most cases teachers are required to continue to accrue credits in order to maintain state certification. The state does not require credits in specific content areas, however, which may explain why the correlation between these predictors is as low as it is. Correlations between each predictor variable and the criterion variable (student pre/post gain) are statistically significant but small. This is not the ideal result for the multiple regression analysis; however, it is consistent with previous research results (e.g., Monk, 1994; Nye et al., 2004; Wayne and Youngs, 2003).

Another test for multicollinearity is the Variance Inflation Factor (VIF) for each predictor. VIF values greater than 10 are cause for concern (Field, 2005, p. 175). For the multiple regression analysis of student variables, VIF values for all predictors fall between 1.03 and 1.67. This confirms that multicollinearity is not an issue for this data set.

It should be noted that the variables related to teacher characteristics represent data from a total of nine teachers; however, the correlation matrix shown in Table 4 was calculated using student data, where each student case includes the values of the teacher-related variables (years of experience, combined coursework credits, and EBAPS score) for that student's teacher. The teacher variables are therefore weighted by the number of participating students assigned to each teacher.

Table 4

Intercorrelations between Predictor Variables and Criterion Variable for Student Multiple Regression Analysis (N = 642)

Variable	WASL	Experience	Coursework	EBAPS
1. S. pre-post gain	.27***	-.16***	-.21***	.12***
2. S. WASL raw score	—	-.09*	-.14***	-.02
3. T. years experience		—	.58***	-.19***
4. T. combined coursework[a]			—	-.35***
5. T. EBAPS score				—

Note. The criterion variable is student gain on the researcher-designed pretests and posttest ("S. pre-post gain"). Variables labeled "S." describe student performance; variables labeled "T." describe teacher characteristics.

[a]Teacher combined coursework ("Coursework" in column heading) is the sum of credits in science and science education; credits in inquiry-oriented teaching strategies; and credits in formative assessment and evaluation of students' prior knowledge.

*$p < .05$, one-tailed. **$p < .01$, one-tailed. ***$p < .001$, one-tailed.

The correlation matrix for the multiple regression analysis of teacher variables is shown in Table 5. The correlation matrix shows that all correlations are small. Correlations between the predictor variables (years of experience and combined coursework credits) and the criterion variable (EBAPS score) are statistically insignificant. The only statistically significant correlation is between the two predictor variables ($r = .45$, $p < .05$). Because regression analysis requires the use of predictors that are correlated to the criterion variable, this lack of statistical significance was considered sufficient to invalidate the results of any regression model, therefore further statistical assumptions were not tested for this data.

Table 5

Intercorrelations between Predictor and Criterion Variables for Teacher Multiple Regression Analysis (N = 15)

Variable	EBAPS	Years experience	Coursework
1. EBAPS score	—	-.20	.03
2. Years experience		—	.45*
3. Combined coursework[a]			—

Note. The criterion variable for the teacher multiple regression analysis is teacher EBAPS score.

[a]Combined coursework ("Coursework" in column heading) is the sum of credits in science and science education; credits in inquiry-oriented teaching strategies; and credits in formative assessment and evaluation of students' prior knowledge.

*$p < .05$, one-tailed.

Absence of Outliers

One assumption of the multiple regression procedure is that the residuals are normally distributed, with no outliers. Outliers in the data can lead to values in the regression equation that do not accurately represent the majority of the data. In order to assess the presence of outliers, casewise diagnostics were run using SPSS. For normally distributed standardized residuals (residuals expressed as *z*-scores), approximately 5% of cases should have an absolute value greater than 2, and approximately 1% of cases should have an absolute value greater than 2.5 (Field, 2005, p. 164). For the multiple regression analysis of student variables, the standardized residual had an absolute value greater than 2 in 31 out of 642 cases, or 4.8%; the standardized residual had an absolute value greater than 2.5 in 5 out of 642 cases, or 0.8%. No cases had standardized residuals with absolute value greater than 3. This confirms the absence of outliers in the data.

Linearity, Normality, Independence, and Homoscedasticity of Residuals

The multiple regression procedure is based on several assumptions. The first is linearity in the relationship between predictors and criterion. Multiple regression produces a linear equation relating the criterion to the predictors; if the data are non-linear, the regression equation will not provide a good description of the relationship. Further assumptions are necessary regarding residuals. For multiple regression analysis to produce generalizable results, the residuals must be normally distributed across the dependent variable. Residuals should also have the same variance at all levels of the predictor variables (homoscedasticity). An additional assumption is independence of errors: the errors in prediction (residuals) should be uncorrelated from case to case (Field, 2005, p. 170).

Figure 1 shows a scatterplot of standardized residuals vs. standardized predicted values of the criterion variable for the multiple regression analysis of student variables. If the relationship between the predictors and the criterion were non-linear, the scatterplot would show a non-random pattern of residuals across predicted values of the criterion. As seen in Figure 1, the residuals in this analysis appear to be approximately randomly and evenly distributed across predicted values, therefore the assumption of linearity was met. This random and even distribution also shows that the residuals had approximately the same variance at all levels of the predictor variables, therefore the assumption of homoscedasticity was also met.

Figures 2 and 3 show a histogram and a normal probability plot, respectively, of the standardized residuals from the multiple regression analysis of student variables. Both figures show that the data met the assumption of normally distributed residuals. The curve superimposed on the histogram in Figure 2 represents the shape of the distribution, and shows that it is a very good approximation of the normal distribution. The diagonal line in Figure 3 represents a normal distribution, and plotted points represent observed standardized residuals; the plotted points fall along the diagonal line, confirming the normality of the standardized residuals.

Independence of residuals (or errors) can be assessed using the Durbin-Watson statistic. This statistic varies between 0 and 4; according to Field (2005), values close to 2 indicate independent residuals (p. 189). The value of the Durbin-Watson statistic for the student multiple regression analysis was 1.9, indicating that the assumption of independent errors was met.

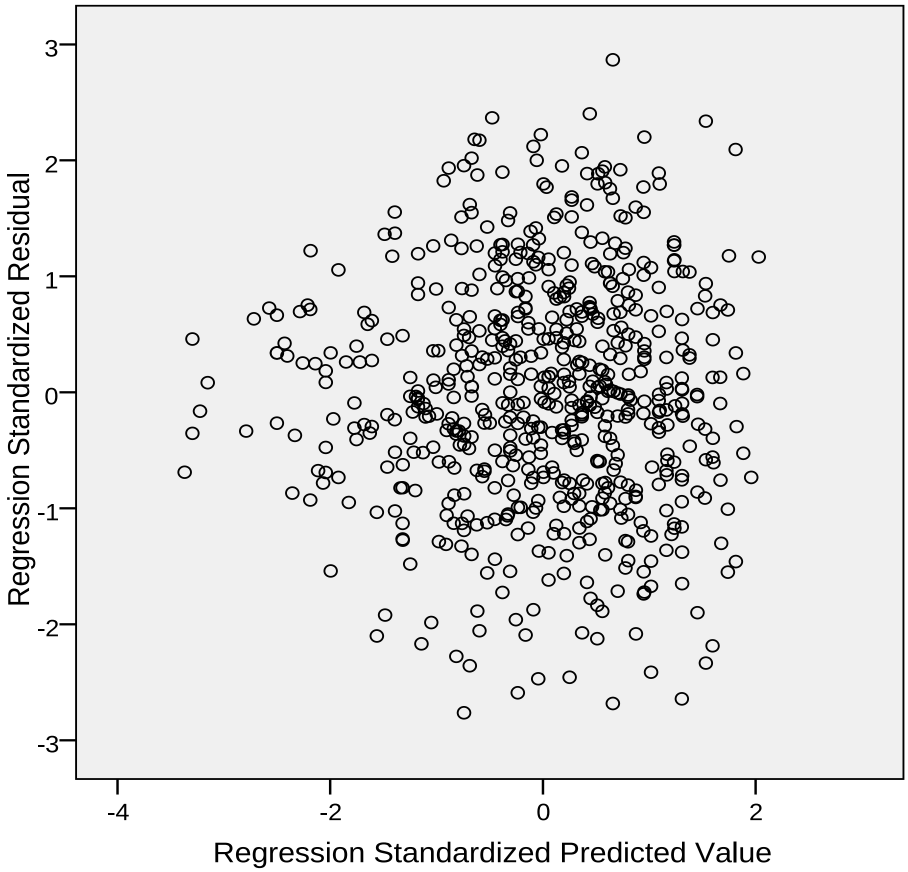


Figure 1. Scatterplot of standardized residuals vs. standardized predicted values for the multiple regression analysis of student variables.

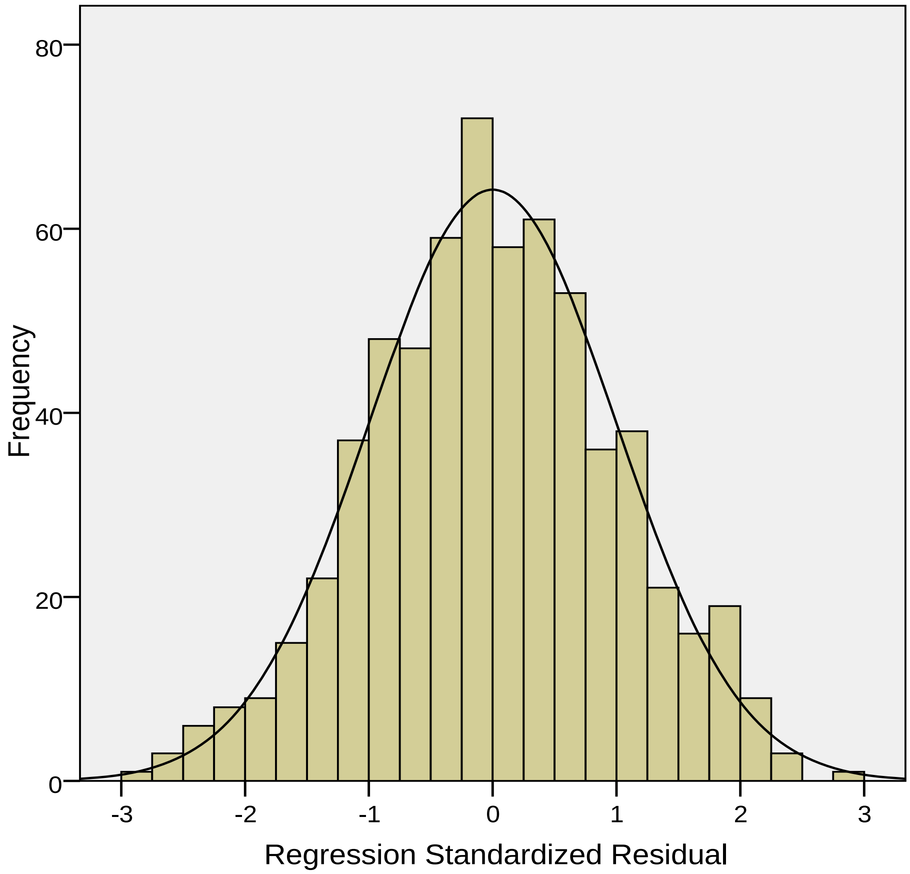


Figure 2. Histogram of standardized residuals from the multiple regression analysis of student variables.

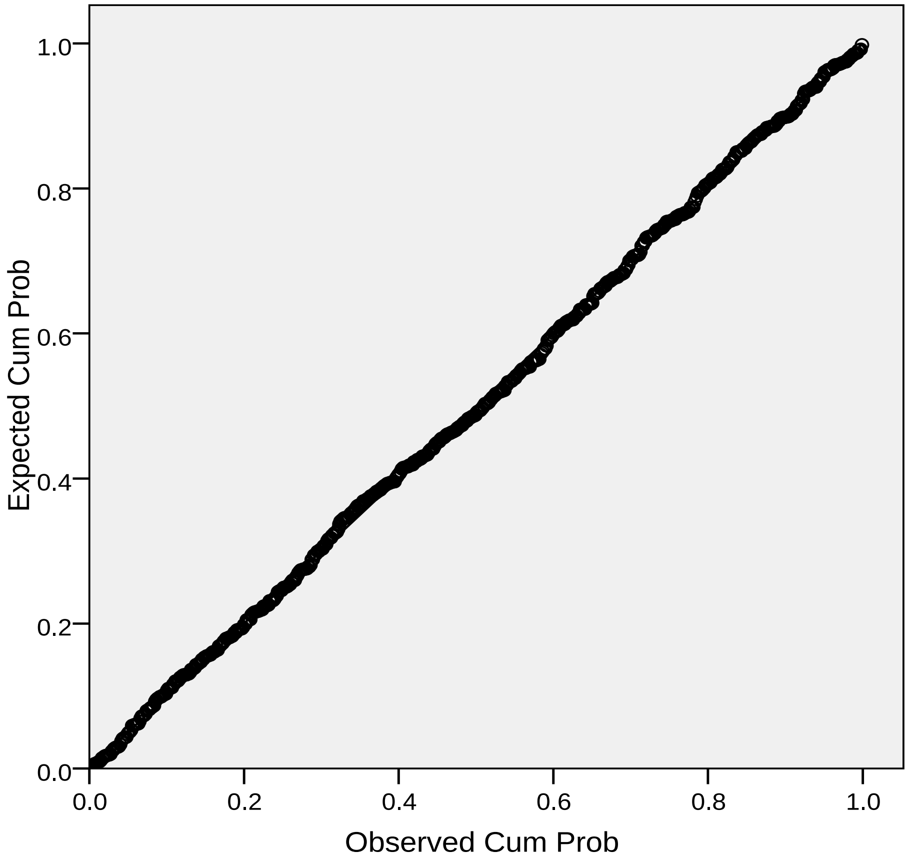


Figure 3. Normal probability plot of standardized residuals in the multiple regression analysis of student variables.

Inferential Statistics: Research Questions 1 and 2

A multiple regression analysis was conducted to predict student gain scores on the researcher-designed assessments from student WASL scores. The results of this analysis indicated that WASL score accounted for a significant amount of the gain variability, $R^2 = .08$, $F(1, 640) = 52.04$, $p < .001$.

A second analysis was conducted to evaluate whether teacher EBAPS score; teacher years of experience; and teacher coursework in science and science education, inquiry-oriented teaching strategies, and formative assessment and evaluation of students' prior knowledge predicted student gain beyond the WASL score. The teacher-related measures accounted for a small but statistically significant proportion of the student gain after controlling for student WASL score, R^2 change = .04, $F(3, 637) = 9.09$, $p < .001$. The full model accounted for approximately 11% of the variance in the criterion, $R^2 = .11$, adjusted $R^2 = .11$. According to Muijs (2004), this indicates a modest fit between the model and the data (p. 166).

The regression model is presented in Table 6, in both unstandardized and standardized forms. While the full model is statistically significant, it should be noted that the regression coefficients for two of the three variables added in Step 2 are not. The variable “teacher combined coursework” is the only variable added in Step 2 with a statistically significant regression coefficient.

Table 6

Summary of Hierarchical Regression Analysis for Variables Predicting Student Pretest-Posttest Gain on Researcher-Developed Content Tests

Variable	*B*	*SE B*	β
Step 1			
Constant	-0.46	0.20	
Student WASL raw score	0.04	0.01	.27***
Step 2			
Constant	-1.87	1.25	
Student WASL raw score	0.03	0.01	.25***
Teacher combined coursework[a]	-0.20	0.08	-.12*

Teacher EBAPS score	0.69	0.36	.08
Teacher years of experience	-0.07	0.06	-.05

Note. R^2 = .08 for Step 1; ΔR^2 = .04 for Step 2 (p < .001).

[a]Teacher combined coursework is the sum of credits in science and science education; credits in inquiry-oriented teaching strategies; and credits in formative assessment and evaluation of students' prior knowledge.

$*p < .05$. $**p < .01$. $***p < .001$.

Inferential Statistics: Research Questions 3 and 4

A multiple regression analysis was conducted to predict teacher EBAPS scores from teacher years of experience. The results of this analysis indicated that teacher years of experience did not account for a significant amount of the variability in teacher EBAPS score, $R^2 = .04$, $F(1, 13) = .54$, $p = .48$. A second analysis was conducted to evaluate whether teacher coursework in science and science education, inquiry-oriented teaching strategies, and formative assessment and evaluation of students' prior knowledge predicted teacher EBAPS score beyond years of experience. The results of this analysis indicated that neither teacher coursework nor teacher coursework combined with years of experience accounted for a significant amount of the variability in teacher EBAPS score, R^2 change = .02, $F(1, 12) = .22$, $p = .65$. As expected from the lack of significant correlations between the predictors and the criterion, neither the Step 1 regression model nor the full regression model was statistically significant.

Summary

Descriptive statistics were computed and reported for all variables. Hierarchical multiple regression analyses were performed in order to address the research questions. The data were examined for violations of underlying assumptions. As a result of this analysis, three variables related to teacher coursework background (credits in science and science education; credits in inquiry-oriented teaching strategies; and credits in formative assessment and evaluation of students' prior knowledge) were combined into one composite variable (teacher combined coursework credits). The hierarchical multiple regression analyses were recalculated and the data were again examined for violations of underlying assumptions. The data used to test research questions 3 and 4 were found to violate the underlying assumptions of multiple regression analysis, as shown in Table 5; therefore, no further analysis was completed on these data.

The data used to test research questions 1 and 2 were examined for outliers and multicollinearity. No outliers were identified, and the variables were judged to have acceptable levels of intercorrelation, as shown in Table 4. The data were further examined for normality and linearity of residuals, homoscedasticity of variance, and independence of error. All conditions were met at acceptable levels, as shown in Figures 1-3; therefore, the multiple regression analysis was accepted as valid and examined for statistical significance. The multiple regression model was found to be statistically significant for both Step 1 and Step 2. The regression model is summarized in Table 6.

Research questions 1 and 2 examine student performance gain on the researcher-designed pretest and posttest. Research question 1 asks whether there is a relationship between student performance gain and scores on the Science WASL. The multiple regression analysis found a statistically significant relationship between student gain and WASL score, $R^2 = .08$, $F(1, 640) = 52.04$, $p < .001$. The revised research question 2 asks whether, beyond the relationship described in research question 1, there is a relationship between student performance gain and the following teacher characteristics: epistemological beliefs, combined coursework, and years of experience. The multiple regression analysis found a statistically significant relationship at this step also, $R^2 = .11$, R^2 change = .04, $F(3, 637) = 9.09$, $p < .001$.This indicates a modest fit between the model and the data (Muijs, 2004, p. 166).

In the full model (Step 2), two of the four regression coefficients were found to be statistically significant: student WASL raw score, and teacher combined coursework. Coefficients for teacher EBAPS score and teacher years of experience were not statistically significant at the $p < .05$ level.

Chapter 5 provides a summary of the purpose, methodology, and findings of this study. The practical significance of the research findings is examined in the context of previously published studies. A discussion of the limitations of the current study is included, and suggestions are made for future research.

Chapter 5: Discussion

Introduction

The purpose of this study was to examine relationships between teacher characteristics and student learning in 8th grade public school science classrooms in western Washington. Two measures were used to quantify student learning: student scores on the science portion of the Washington Assessment of Student Learning (WASL), and student gain (posttest score minus pretest score) on researcher-created content tests administered in the science classes. Two instruments were used to collect information about teacher characteristics: a Teacher Questionnaire incorporating questions about teacher coursework background, professional development, and school and administrative support (see Appendix A); and the Epistemological Beliefs Assessment for Physical Science (EBAPS) (see Appendix B).

This chapter provides a summary of the purpose, methodology, and findings of this study. The practical significance of the research findings is examined in the context of previously published studies. A discussion of the limitations of the current study is included, and suggestions are made for future research.

Purpose of the Study and Research Methodology

The current science education reform movement identifies the kinds of science content understanding and conceptions about the nature of science students should develop during their K-12 education. However, there is a shortage of empirical evidence on what kinds of teacher education and professional development lead to the desired student learning. While some principles of effective teaching are described in the science education reform literature, there is a scarcity of rigorous research on the effects of teacher preparation and professional development on teaching practice and student learning.

The question of how best to prepare and support K-12 science teachers for reformed teaching is a critical and unresolved question. As described in the research review in chapter 2, many intermediate steps have been examined and documented; however, the critical link between teacher characteristics and student learning in science is not well studied. This study was informed by existing research relating teacher characteristics to student learning, as well as research examining other relevant teaching and learning relationships, and was intended to contribute to the knowledge base for the design of effective professional development for teachers of science.

Review of Results

A summary of study findings is presented in order of the research questions posed in chapter 1. Complete study results, with relevant tables and figures, are presented in chapter 4.

Research Questions 1 and 2

The first research question asked if there is relationship between student gain on the researcher-designed content tests and student scores on the Science WASL. There was a statistically significant relationship between student gain score and student score on the Science WASL. As shown by the correlation matrix, there is a positive correlation between student gain and Science WASL score, $r = .27$, $p < .001$. As shown by the regression model, Science WASL score accounted for a significant amount of the gain score variability, $R^2 = .08$, $F(1, 640) = 52.04$, $p < .001$. According to this model, approximately 8% of the variance in student gain score was accounted for by differences in student score on the Science WASL.

The second research question asked if, beyond any relationship between students' gain scores on the researcher-designed content tests and their scores on the Science WASL, there is a relationship between students' gain on the researcher-designed content tests and teachers' epistemological beliefs, coursework in inquiry-oriented teaching strategies, coursework in formative assessment and evaluation of students' prior knowledge, coursework in science content or science education, or years of experience. After initial data analysis showed problematic multicollinearity between teacher coursework variables, this research question was revised to ask if, beyond the relationship described in research question 1, there is a relationship between student performance gain and the following teacher characteristics: epistemological beliefs, combined coursework, and years of experience. Combined coursework is the sum of teacher coursework credits in the areas of science content, science education, inquiry-oriented teaching strategies, and formative assessment and evaluation of students' prior knowledge, as reported on the Teacher Questionnaire.

The second step of the hierarchical multiple regression analysis showed that there is a statistically significant relationship between student gain score and the teacher variables described above after controlling for student Science WASL score, R^2 change = .04, $F(3, 637) = 9.09$, $p < .001$. The measured teacher characteristics accounted for approximately 4% of variance in student gain scores after controlling for students cores on the Science WASL. The full regression model accounted for approximately 11% of the variance in student gain, $R^2 = .11$, adjusted $R^2 = .11$. Both teacher years of experience and teacher combined coursework were negatively correlated with student gain score ($r = -.16$, $p < .001$, and $r = -.21$, $p < .001$, respectively); these predictors also had negative regression coefficients. Teacher EBAPS score was positively correlated with student gain score, $r = .12$, $p < .001$; the regression coefficient for teacher EBAPS score was also positive.

Research Questions 3 and 4

The third research question asked if there is a relationship between teachers' epistemological beliefs, as measured by the EBAPS, and their years of teaching experience. The correlation matrix showed that the correlation between teachers' EBAPS score and years of experience is not statistically significant for this sample. The regression model was also not statistically significant.

The fourth research question asked if, beyond any relationship between teachers' epistemological beliefs and their years of teaching experience, there is a relationship between teachers' epistemological beliefs and their coursework in science content or science education, in inquiry-oriented teaching strategies, or in formative assessment and evaluation of students' prior knowledge. After initial data analysis showed problematic multicollinearity between teacher coursework variables, this research question was revised to ask if, beyond the relationship described in research question 3, there is a relationship between teachers' epistemological beliefs and their combined coursework. For this research question also, both the correlation matrix and the regression model failed to show statistically significant relationships.

Discussion of Results: Research Questions 1 and 2

In the analysis for research questions 1 and 2, student score on the Science WASL was used as a statistical control representing overall student achievement. As noted by Wayne and Youngs (2003), when looking for meaningful relationships between teacher characteristics and student learning it is important to control for student background characteristics in some way (p. 92). Wayne and Youngs looked for studies that included socioeconomic status as a control for student background characteristics. In this study, student score on the Science WASL was used for this purpose.

As described in chapter 3, the Science WASL addressed all areas of science, including inquiry in science and applications of science. Because the researcher-designed content pretest and posttest were specific to a single physical science unit, and the Science WASL was a much more general assessment addressing content covered over several years, student performance on the 8th grade Science WASL might be expected to be only moderately influenced by the characteristics of the 8th grade science teacher. The Science WASL score was therefore used to statistically control for the effects of general student characteristics. It is important to note that the Science WASL was administered after the relevant unit of instruction (including pretest and posttest) was competed; therefore, to the extent that student learning gain at the end of the science unit in winter was retained through the end of the school year, any overlap in the content of the Science WASL with the researcher-designed pretest and posttest could be expected to increase the correlation between student learning gain and Science WASL score, thus decreasing the degree to which Science WASL score could serve as an effective statistical control for overall student achievement. However, according to the regression model only 8% of the variance in student gain score was accounted for by differences in student score on the Science WASL. This suggests limited overlap in the student learning measured by the Science WASL as compared to the researcher-designed pretest and posttest, which supports the use of Science WASL score as a measure of overall student achievement.

The second step of the hierarchical multiple regression analysis looked for relationships between student learning gain and teacher characteristics, beyond the student gain accounted for by variability in student score on the Science WASL. Consistent with previous research, the present study found only limited relationships between student learning and teacher characteristics, and some relationships were in the opposite direction of what might be hoped for. In their systematic review of research on teacher characteristics and student learning gains, Wayne and Youngs (2003) found that most studies of the effects of teacher degrees and coursework found indeterminate results; among studies that did find statistically significant relationships, some were found to be positive and some negative (p. 101). The present study found positive relationships between teacher EBAPS score and student gain, and negative relationships between student gain and teacher years of experience and coursework in science, science education, inquiry teaching strategies, and formative assessment.

Wayne and Youngs (2003) suggested that some of the indeterminate or conflicting results in the literature are at least partly explained by the absence of subject-related information (p. 101). This is consistent with Monk's 1994 study, which looked at teacher coursework broken down by subject area and found complex relationships between teacher coursework and student learning. In most cases, more undergraduate mathematics or science courses taken by teachers correlated with a small increase in student mathematics or science performance (pp. 130-137). However, undergraduate life science preparation had a negative effect on student science achievement for juniors (p. 136); also for juniors, graduate coursework in life science showed a positive effect on science achievement, but graduate coursework in physical science showed a negative effect (p. 136). Monk also found interaction effects between undergraduate coursework and subject taught. For example, the relationship between teacher undergraduate coursework in physical science and student achievement was positive for sophomores enrolled in life science classes, but negative for sophomores enrolled in physical science (p. 141).

As shown by the research described above and in chapter 2, the relationship between teacher coursework and student gain is complex and subject-specific. Due to the small number of teachers in this study and the intercorrelations between teacher responses to the various categories of coursework described on the Teacher Questionnaire, it was not possible in this study to differentiate between coursework in different areas of science, between coursework in science and coursework in science education, or between different emphases in science education coursework (e.g., inquiry-oriented teaching methods, formative assessment strategies). Therefore, the negative relationship found in this study between student gain and the combined teacher coursework variable should be interpreted not as a sign that teacher coursework harms student learning, but rather as an indication that the data entered into the analysis was not sufficiently specific. This result is consistent with a number of studies that have found that, while teachers have a large influence on student learning, identifying specific teacher characteristics responsible for the influence is difficult (e.g., Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Wright, Horn, & Sanders, 1997).

As described in chapter 4, in the present study complete teacher data with matched student data was available for nine teachers. The small number of teachers increases the possibility of sample-specific findings. For example, the most experienced teacher included in this analysis had a significant adverse health event during the time of instruction for the science unit in question, and was absent from the classroom for some time. It is difficult to assess the impact of this event on student learning, and the impact of differences in student learning caused by this event on the present study.

One of the teacher characteristics examined in the present study, teacher EBAPS score, is a measure of teacher epistemological beliefs in the context of physical science, including questions that probe teacher understanding of the nature of science. Of the studies discussed above and in chapter 2 relating teacher characteristics to student learning, none included measures of teacher epistemological beliefs or understanding of the nature of science. This study therefore extends the existing research slightly by examining the relationship between teacher epistemological sophistication and student learning. Though teacher EBAPS scores in this study had limited range (from 2.95 to 3.57 on a scale of 0 to 4), EBAPS score was positively correlated with student gain score, $r = .12$, $p < .001$. The regression coefficient for teacher EBAPS score was also positive.

Some existing studies have examined relationships between teacher epistemological beliefs and teacher classroom practice, with inconclusive results: some studies supported a direct influence of teachers' understanding of the NOS on their classroom practice (e.g., Brickhouse, 1989), and other studies supported the position that there is no influence (e.g., Lederman & Zeidler, 1987; Mellado, 1997). Other study results have suggested that a sophisticated understanding of the nature of science is a necessary but not sufficient condition for teacher classroom practice consistent with the nature of science (e.g., Schwartz and Lederman, 2002; Tobin & McRobbie, 1997). Though the present study did not examine classroom practice directly, the use of Science WASL as a statistical control for student overall achievement makes it likely that correlations between teacher characteristics and student learning are related to teacher practice. Therefore, the results of this investigation provide some support for the findings of previous studies linking teacher epistemological beliefs with classroom practice. As stated above, however, relationships between teacher characteristics and student learning found in the present analysis must be interpreted with caution due to the small number of teachers included in the analysis.

Discussion of Results: Research Questions 3 and 4

In the analysis for research questions 3 and 4, correlations between the predictors (teacher years of experience and combined coursework credits) and the criterion variable (teacher EBAPS score) were found to be small and statistically insignificant. The only statistically significant correlation found was between the two predictor variables ($r = .45$, $p < .05$). The regression model also failed to show statistically significant relationships between the predictors and the criterion.

This result is consistent with previous research. Several previous studies of teachers' understanding of the nature of science have found that teachers' understanding did not seem to be affected by their years of teaching experience (e.g., Carey & Stauss, 1970; Kimball, 1967; Miller, 1963) or by the number of courses they had taken (e.g., Billeh & Hasan, 1975; Carey & Stauss, 1968, 1969, 1970). Some research has suggested that targeted coursework can improve teacher understanding of the nature of science (e.g., Akerson, Abd-El-Khalick, & Lederman, 2000; Akerson, Morrison, & McDuffie, 2006; Billeh & Hasan, 1975; Carey & Stauss, 1968, 1970); however, the data used in the present study did not allow for differentiation between coursework targeted at epistemological issues and more general science or science education coursework.

Two categories of coursework were originally included with the intent of examining more specific relationships with teacher EBAPS score; these categories were coursework in inquiry-oriented teaching strategies and coursework in formative assessment and evaluation of students' prior knowledge. However, initial analysis found unacceptable levels of intercorrelation between these coursework categories and the third coursework category, coursework in science and science education. The three variables related to teacher coursework were therefore combined into one combined coursework variable, removing the possibility of examining relationships between teacher EBAPS score and coursework in particular areas.

Limitations of the Study

Design

The design of this study was correlational. In correlational research, it cannot be assumed that relationships found in the data are causal. A significant correlation coefficient between two variables could be due to a causal relationship; if so, no indication of the direction of causality is provided by the correlation coefficient. Alternatively, the correlation could be due to a third variable that is causally related to both of the measured variables (Field, 2005, p. 128). Given the complexity of the relationships found in previous research on student learning (e.g., Monk, 1994; Wayne & Youngs, 2003), it is particularly important to keep this limitation in mind when considering the results of research on teachers and student learning, including the present study.

Sampling

The greatest limitation of the present study is the use of convenience sampling. Because a convenience sample is not necessarily representative of the larger population, the results of the analysis may not be generalizable to a larger population (Gall, Gall, & Borg, 2003, p.175). Teachers approached for the study were free to decline to participate, either by responding in the negative or by simply choosing not to respond to the request for participation. Since teachers had access to both the Teacher Questionnaire and the EBAPS when making the choice to participate or decline, it is possible that the teachers who chose to decline did so because they were uncomfortable with the questions included in the instruments. This differential self-selection is a threat to the external validity of the study.

The sample size presents an additional limitation to the present analysis. Tabachnick and Fidell (2007) have suggested that a sample size exceeding 100 cases is necessary for testing individual predictors in a multiple regression analysis (p. 123). The number of student cases included in the analysis is much greater than 100, but the number of teacher cases is far fewer. While research questions 1 and 2 used matching teacher and student data and a total of 642 cases were entered into the analysis, only nine teachers were represented in these cases. The analysis of research questions 3 and 4 included teacher data consisting of a total of 15 cases, far fewer than the recommended number.

This small sample size may have contributed to that lack of statistically significant findings in the analysis of relationships among teacher variables.

Methodology

One methodological concern is the use of self-reported data. Most of the teacher characteristics included in the present analysis were measured using the Teacher Questionnaire, a self-report instrument administered through an online survey tool. The use of an online questionnaire did not permit clarification of questionnaire items. Teacher reports of coursework credits were not verified through independent sources such as transcripts. Some items on the Teacher Questionnaire asked teachers to report the number of coursework credits in various areas before the time they began full-time teaching, as well as credits taken after that time. Since many of the teachers in the study reported having between 5 and 20 years of teaching experience, these teachers were reporting credits taken a decade or more in the past, and credit totals may not have been completely accurate.

A second methodological concern is the correct specification of predictor variables. The use of multiple regression analysis carries with it the risk of specification errors. Regression analysis assumes that all relevant predictors are included in the model, and no irrelevant predictors are present (Licht, 1995, p. 49). According to Licht, "the values of all indexes from [multiple regression and correlation analysis] can change dramatically when even one important predictor is added to the analysis" (p. 49). This means that failure to include an important predictor variable can result in misleading regression results. Unfortunately, the task of identifying all variables relevant to student learning gain has proven difficult at best, even if only teacher-related variables are considered. This is exemplified in the results of studies described in chapter 2 that found a large teacher influence on student learning, but failed to find specific teacher variables to account for the large teacher effect (e.g., Nye et al., 2004; Rivkin et al., 2005; Wright et al., 1997). Given this difficulty, the relationships found in the present study should be considered suggestive and further work should be done to identify relevant teacher variables.

Data

The data used in the present analysis present several limitations to the study. As stated above, the number of complete teacher cases ($N = 15$) was much lower than the number of cases recommended for multiple regression analysis. This could explain the lack of statistically significant findings in the analysis for research questions 3 and 4. In addition, the range of teacher scores on the EBAPS was somewhat narrow: the minimum and maximum scores were 2.95 and 3.58 (3.57 for teachers included in the analysis of student learning gain), respectively, out of a possible range of 0 to 4. A score of 4 on the EBAPS represents sophisticated epistemology, while a score of 0 represents naïve or unsophisticated epistemology. The teacher scores ($M = 3.25$, $SD = 0.19$) indicate that the teachers who completed the EBAPS for this study were, on average, epistemologically more sophisticated than naïve. A number of previous studies have found that teachers are not, on average, epistemologically sophisticated, at least with respect to understanding of the nature of science (see, e.g., Akerson, Abd-El-Khalick, & Lederman, 2000; Akerson, Morrison, & McDuffie, 2006; Carey & Stauss, 1968, 1969, 1970). This suggests that the teachers who participated in the present study may not be representative of the overall teacher population. Previous studies, however, used different instrumentation from the present study; it is difficult to assess the degree to which the teacher data used in this study is representative of a larger population without comparative EBAPS data from a larger sample of teachers.

An additional limitation to the data is the small average student gain score (M = 1.00, SD = 1.42). Many students had zero or negative gain from pretest to posttest. This may be due to a difference in emphasis between the researcher-designed tests and the curriculum used to teach the unit. Though the concepts assessed by the pretest and posttest were addressed in the curriculum, and in many cases represented foundational concepts for the subject area of the curriculum, participating teachers reported that some of the concepts tested were only minimally addressed in the curriculum materials. It is possible that relationships between teacher characteristics and student learning would be different if the curriculum were better aligned with the assessments, or vice versa. It should be noted that teachers participating in the present study used the results of the researcher-designed assessments to inform their instruction of the same unit the following year, choosing to supplement the curriculum in the area of several foundational concepts that were minimally addressed in the curriculum and for which student learning gains were found to be unacceptably low.

Implications of the Findings and Suggestions for Future Research

Consistent with many previous studies, the present study provides modest support for the notion that student learning is not strongly related to teachers' years of experience or to the number of courses teachers have taken. This suggests that efforts to improve science teacher preparation must go beyond requiring more science coursework. Further research is needed to identify whether more targeted teacher coursework leads to improvements in student learning. The present study originally included more detailed information about particular categories of teacher coursework; however, intercorrelations between categories necessitated combining them into one larger, less specific category. Future studies could avoid this outcome by carefully defining coursework categories to be mutually exclusive. Future studies must also include a larger number of teacher cases in order to make the use of multiple regression analysis statistically sound and increase the likelihood of statistically significant findings.

As discussed above, in the present study the alignment between researcher-designed assessments and the science curriculum used by participating teachers was imperfect. Further research conducted in settings with higher alignment between assessment and curriculum could add to our understanding of the relationships between teacher characteristics and student learning. Were such a setting to additionally include teacher professional development aligned with the curriculum and with student learning goals, research conducted in this setting could also assess the relationship between student learning and targeted teacher coursework, rather than the more general teacher coursework measured in the present study. Such a study could fill a gap in the currently existing research.

The present study provides preliminary evidence suggesting that teacher epistemology is positively correlated with student learning. Previous studies have not examined this relationship. The result of this analysis is consistent with the finding of some previous studies that sophisticated teacher understanding of the nature of science is necessary (though not sufficient) for teaching practice consistent with the nature of science (e.g., Brickhouse, 1989; Schwartz and Lederman, 2002; Yerrick, Parke, &

Nugent, 1997); and with other studies that have found such teaching practice to be correlated with increased student learning (e.g., Adamson et al., 2003; Marx et al., 2004; Sawada et al., 2002; Schroeder, Scott, Tolson, Huang, & Lee, 2007). While the relationship found in the present analysis is small, it suggests an area for further research. Teachers participating in the present study were found to have a rather narrow range of scores on the EBPAS; in order to explore this relationship more fully, future research in this area should include teachers with a wider range of epistemological sophistication.

Concluding Remarks

The current science education reform movement in the United States began in response to the warning in *A Nation at Risk* (National Commission on Excellence in Education, 1983) that our society was threatened by “a rising tide of mediocrity” in our education system. The most influential publications in this movement for reformed science teaching have been those of Project 2061, a project of the American Association for the Advancement of Science (AAAS), and the *National Science Education Standards*, published by the National Research Council (1996). The goal of this reform effort is science literacy for all high school graduates.

The final test of any science education reform effort is its impact on the understanding of science students. However, as described in chapter 2, there is a scarcity of rigorous research on the effects of teacher preparation and professional development on teaching practice and student learning. Little empirical research has been done that directly relates teacher characteristics to student learning, and those studies that have been done have found inconclusive or mixed results (Greenwald, Hedges, & Laine, 1996; Monk, 1994; Wayne & Youngs, 2003; Wilson, Floden, & Ferrini-Mundy, 2002). This study is designed to contribute to the research base relating teacher characteristics to student learning.

Most previous studies of this kind have used general measures of student learning such as scores on the National Assessment of Educational Progress (NEAP; Monk, 1994) or other standardized exams at the state or national level (Goldhaber and Brewer, 1997; Greenwald, Hedges, & Laine, 1996), rather than measures specific to the content addressed in the classrooms of the teachers and students from whom data are gathered. The present study used more curriculum-specific content tests designed by the researchers in order to more accurately assess the relationship between teacher characteristics and student learning in the content addressed in the classroom. The results of this analysis showed that at least in the present case, student learning gain in content addressed in the classroom was not strongly correlated to student performance on the state standardized science assessment. This suggests that research seeking relationships between teacher characteristics and student learning may be more effective if researchers use more targeted content assessments.

Despite the use of targeted content assessments, the present study did not find strong relationships between student learning gains and any of the measured teacher variables. As discussed above, this may be due in part to specification error: multiple regression analysis assumes that all relevant predictor variables are measured, but in the domain of student learning gains, research has not identified a comprehensive set of predictor variables, therefore this assumption is violated. It is also likely that insufficient specificity in measured teacher background variables contributed to the weakness of the relationships found in this study. The results of several previous studies suggest that undifferentiated counts of teacher coursework credits are not likely to be strongly related to student learning (e.g., Wayne & Youngs, 2003), whereas more specific coursework information has provided at least some insight into such relationships (e.g., Monk, 1994). The present study was intended to examine more specific teacher coursework categories; however, this specificity was lost when data categories were combined because of multicollinearity between predictor variables in the initial analysis.

While previous studies have suggested the existence of a relationship between teacher epistemology and student learning, as described in chapter 2, no study has examined the relationship directly. The present study attempted to address this shortcoming in the literature. The results suggest a small positive relationship between teacher epistemological sophistication and student learning. Further research with a larger number of teachers is needed to explore this relationship.

As noted in *Blueprints for Reform* (AAAS, 1998) and again by Wilson, Floden, and Ferrini-Mundy (2002) in their national review of research on teacher preparation, there is shortage of rigorous research on the effects of either subject matter knowledge preparation or pedagogical preparation on teacher knowledge and student learning. While there is general agreement, supported by research results, that teachers play a critical role in helping students attain science literacy, there is no similar consensus about "what it means to be an effective teacher" and what kind of preparation contributes to effective teaching (AAAS, 1998, p. 82). The present study was intended to contribute to this critical research area. The findings of this study do not fill the research gaps described above; however, they suggest directions for future study and contribute to the body of relevant research, working toward the ultimate goal of improving our knowledge of effective teacher preparation and thereby enhancing student learning of science.

References

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

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics – Physics Education Research, 2,* 010101.

Adamson, S., Banks, D., Burtch, M., Cox, F., III, Judson, E., Turley, J., et al. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching, 40,* 939-957.

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

Akerson, V. L., Morrison, J. A., & McDuffie, A. R. (2006). One course is not enough: Preservice elementary teachers' retention of improved views of nature of science. *Journal of Research in Science Teaching, 43,* 194-213.

American Association for the Advancement of Science. (1989). *Science for all Americans.* New York: Oxford University Press.

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

American Association for the Advancement of Science. (1998). *Blueprints for reform.* New York: Oxford University Press.

Billeh, V. Y., & Hasan, O. E. (1975). Factors affecting teachers' gain in understanding the nature of science. *Journal of Research in Science Teaching, 12,* 209-219.

Brickhouse, N. W. (1989). The teaching of the philosophy of science in secondary classrooms: Case studies of teachers' personal theories. *International Journal of Science Education, 11,* 437-449.

Bruner, J. (1961). *The process of education.* Cambridge, MA: Harvard University Press.

Bruner, J. (1962). *On knowing: Essays for the left hand.* Cambridge, MA: The Belknap Press of Harvard University Press.

Bruner, J. (1966). *Toward a theory of instruction.* Cambridge, MA: The Belknap Press of Harvard University Press.

Bruner, J. (1996). *The culture of education.* Cambridge, MA: Harvard University Press.

Carey, R. L., & Stauss, N. G. (1968). An analysis of the understanding of the nature of science by prospective secondary science teachers. *Science Education, 52,* 358-363.

Carey, R. L., & Stauss, N. G. (1969). An analysis of the relationship between prospective science teachers' understanding of the nature of science and certain academic variables. *Bulletin of the Georgia Academy of Science, 27*(3), 148-158.

Carey, R. L., & Stauss, N. G. (1970). An analysis of experienced science teachers' understanding of the nature of science. *School Science and Math, 70,* 366-376.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12,* 311-329.

Committee on Prospering in the Global Economy of the 21st Century & Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: National Academies Press. Retrieved July 11, 2008, from http://books.nap.edu/openbook.php?record_id=11463&page=1

Dewey, J. (1915). *The school and society* (Rev. ed.). Chicago: The University of Chicago Press.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education.* New York: The Macmillan Company.

Dewey, J. (1938). *Experience and education.* New York: Simon & Schuster.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (Eds.) (1999). *How people learn: Bridging research and practice.* Washington, DC: National Academy Press.

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from *High school and beyond*. *Economics of Education Review, 13*(1), 1-17.

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

Ellis, A. K. (2001). *Research on educational innovations* (3rd ed.). Larchmont, NY: Eye On Education.

Epistemological beliefs assessment for physical science (EBAPS). (n.d.). Retrieved August 26, 2008, from http://www2.physics.umd.edu/~elby/EBAPS/home.htm

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives, 7*(9). Retrieved July 5, 2008, from http://epaa.asu.edu/epaa/v7n9.html

Field, A. (2005). *Discovering statistics using SPSS: (and sex, drugs and rock 'n' roll)* (2nd ed.). London: Sage Publications.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn and Bacon.

Gallagher, J. J. (1991). Prospective and practicing secondary school science teachers' knowledge and beliefs about the philosophy of science. *Science Education, 75,* 121-133.

Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In W. Fowler, Jr. (Ed.), *Developments in school finance, 1996* (pp. 197-210). Washington, DC: National Center for Educational Statistics, U.S. Department of Education.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66,* 361-396.

Grossman, P. L., & Stodolsky, S. S. (1995). Content as context: The role of school subjects in secondary school teaching. *Educational Researcher, 24*(8), 5-11 & 23.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66,* 64-74.

Hestenes, D., Wells, M., & Swackhammer, G. (1992). Force concept inventory. *The Physics Teacher, 30,* 141-151.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67,* 88-140.

Horner, J. K., & Rubba, P. (1978). The myth of absolute truth. *The Science Teacher, 45*(1), 29-30.

Horner, J. K., & Rubba, P. (1979). The laws are mature theories fable. *The Science Teacher, 46*(2), 31.

The idea behind the EBAPS. (n.d.). Retrieved August 26, 2008, from http://www2.physics.umd.edu/~elby/EBAPS/idea.htm

Judson, E., & Sawada, D. (2001). *Tracking transfer of reform: Tracking transfer of reform methodology from science and math college courses to the teaching style of beginning teachers of grades 5-12.* (Report No. ACEPT-TR-PRG01-2). Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. (ERIC Document Reproduction Service No. ED455208)

Kim, E., & Pak, S. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics, 70,* 759-765.

Kimball, M. E. (1967). Understanding the nature of science: A comparison of scientists and science teachers. *Journal of Research in Science Teaching, 5,* 110-120.

Lawson, A., Benford, R., Bloom, I., Carlson, M., Falconer, K., Hestenes, D., et al. (2002). Evaluating college science and mathematics instruction: A reform effort that improves teaching skills. *Journal of College Science Teaching, 31,* 388-393.

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

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

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

Levin, H. M. (1970). A new model of school effectiveness. In *Do teachers make a difference? A report on recent research on pupil achievement* (pp. 55-78). Washington, DC: U.S. Government Printing Office.

Lewis, J., & Ouellette, H. (1979). Teacher training, teacher experience and pupil achievement. *Journal of Instructional Psychology, 6*(4), 3-5.

Licht, M. H. (1995). Multiple regression and correlation. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 19-64). Washington, DC: American Psychological Association.

Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics, 73,* 372-382.

Mackay, L. D. (1971). Development of understanding about the nature of science. *Journal of Research in Science Teaching, 8,* 57-66.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Solowy, E., Geier, R., et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41,* 1063-1080.

Matthews, D. R., & McLaughlin, T. F. (1994). Effects of learner-centered laboratory activities on achievement and students' preferences in two high school biology courses. *Perceptual and Motor Skills, 78,* 285-286.

Mazur, E. (1997). *Peer instruction: A user's manual.* Upper Saddle River, New Jersey: Prentice Hall.

McDermott, L. C., & Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity, Part I: Investigation of student understanding. *American Journal of Physics, 60,* 994-1003.

McDermott, L. C., Shaffer, P. S., & Constantinou, C. P. (2000). Preparing teachers to teach physics and physical science by inquiry. *Physics Education, 35,* 411-416.

McDermott, L. C., & University of Washington Physics Education Group. (1996). *Physics by Inquiry*. New York: John Wiley.

Mead, M., & Métraux, R. (1957). Image of the scientist among high-school students: A pilot study. *Science, 126,* 384-390.

Mellado, V. (1997). Preservice teachers' classroom practice and their conceptions of the nature of science. *Science and Education, 6,* 331-354.

Michelson, S. (1970). The association of teacher resourceness with children's characteristics. In *Do teachers make a difference? A report on recent research on pupil achievement* (pp. 120-168). Washington, DC: U.S. Government Printing Office.

Miller, P. E. (1963). A comparison of the abilities of secondary teachers and students of biology to understand science. *Iowa Academy of Science, 70,* 510-513.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects on pupil performance in secondary mathematics and science: The case of teacher subject-matter preparation. In R. G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29-58). Ithaca, NY: ILR Press.

Muijs, D. (2004). *Doing quantitative research in education with SPSS.* London: Sage Publications.

National Academy of Sciences. (2000). *Science and technology concepts for middle schools: Properties of matter.* Burlington, NC: Carolina Biological Supply Company.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

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

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237-257.

Office of Superintendent of Public Instruction. (2006). *Frequently asked questions about scoring the Washington assessment of student learning (WASL).* Retrieved August 28, 2008, from http://www.k12.wa.us/Assessment/pubdocs/ScoringtheWASL_FAQ100406.pdf

Osman, M. E., & Hannafin, M. J. (1994). Effects of advance questioning and prior knowledge on science learning. *Journal of Educational Research, 88,* 5-13.

Pearson Educational Measurement. (2007). *Washington Assessment of Student Learning: Grade 8 2006 technical report.* Retrieved August 28, 2008, from http://www.k12.wa.us/assessment/TechRpts/TechReport2006/2006Gr8WASLTechReport.pdf

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417-458.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., et al. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102,* 245-253.

Schmidt, D. J. (1967). Test on understanding science: A comparison among several groups. *Journal of Research in Science Teaching, 5,* 355-356.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44,* 1436-1460.

Schwartz, R., & Lederman, N. (2002). “It’s the nature of the beast”: The influence of knowledge and intentions on learning and teaching nature of science. *Journal of Research in Science Teaching, 39,* 205-236.

Science Assessment Team of the Washington Office of the Superintendent of Public Instruction. (2007). *Washington Assessment of Student Learning: Science specifications, grades 5, 8, and 10, 2007-08.* Retrieved August 28, 2008, from http://www.k12.wa.us/assessment/WASL/Science/WASLSpecifications.aspx

Shapiro, B. L. (1996). A case study of change in elementary student teacher thinking during an independent investigation in science: Learning about the "face of science that does not yet know." *Science Education, 80,* 535-560.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15(*2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Shulman, L. S. (2002). Truth *and* consequences? Inquiry and policy in research on teacher education. *Journal of Teacher Education, 53,* 248-253.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5*(1), 41-48.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn and Bacon.

Tobin, K., & McRobbie, C. (1997). Beliefs about the nature of science and the enacted science curriculum. *Science and Education, 6,* 355-371.

WASL Timelines and Calendars. (n.d.) Retrieved August 26, 2008, from http://www.k12.wa.us/assessment/WASL/Timelines-Calendars.aspx

Washington Assessment of Student Learning (WASL). (n.d.). Retrieved August 26, 2008, from http://www.k12.wa.us/assessment/WASL/default.aspx

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73,* 89-122.

Which EBAPS items belong to which subscales? (n.d.). Retrieved August 28, 2008, from http://www2.physics.umd.edu/~elby/EBAPS/EBAPS_subscale_sort.doc

Wilson, S. M., Floden, R. E., & Ferrini-Mundi, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education, 53,* 190-204.

Wise, K. C. (1996). Strategies for teaching science: What works? *The Clearing House, 69,* 337-338.

Wosilait, K., Heron, P. R. L., Shaffer, P. S., & McDermott, L. C. (1999). Addressing student difficulties in applying a wave model to the interference and diffraction of light. *Physics Education Research, American Journal of Physics Supplement, 67* (7) S5.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-67.

Yerrick, R., Parke, H., & Nugent, J. (1997). Struggling to promote deeply rooted change: The "filtering effect" of teachers' beliefs on understanding transformational views of teaching science. *Science Education, 81,* 137-159.

Appendix A: Teacher Questionnaire

*1. School district:

*2. ID Number

*3. Select the grade level at which you teach:

☐ Elementary School ☐ Middle School ☐ High School

*4. Part I-A: Professional experience

	More than 20 years	10 to 20 years	5 to 9 years	2 to 4 years	1 year or less
a. How many years have you been teaching?	○	○	○	○	○
b. How long at your current grade level?	○	○	○	○	○
c. How long have you been in this district?	○	○	○	○	○

*5. Part I-B: Coursework background -- List courses taken *before* you began full-time teaching. Include undergraduate and graduate courses. (Note: typically one credit-hour corresponds to 10 hours of class time)

	0	1-3	4-9	10-15	16-30	more than 30
a. How many credit-hours of mathematics coursework have you taken?	○	○	○	○	○	○
b. How many credit-hours of physical science coursework (e.g., physics, chemistry, geology) have you taken?	○	○	○	○	○	○

	0	1-3	4-9	10-15	16-30	more than 30
c. How many credit-hours of life science coursework (e.g., biology) have you taken?	○	○	○	○	○	○
d. How many credit-hours of mathematics education coursework have you taken?	○	○	○	○	○	○
e. How many credit-hours of science education coursework have you taken?	○	○	○	○	○	○
f. How many credit-hours of coursework have you taken that focused on the use of questioning, discussion, and inquiry-oriented teaching strategies?	○	○	○	○	○	○
g. How many credit-hours of coursework have you taken that focused on using technology in the classroom?	○	○	○	○	○	○
h. How many credit-hours of coursework have you taken that focused on the use of formative assessment and evaluation of students' prior knowledge to inform instruction?	○	○	○	○	○	○

*6. Part I-B: Coursework background (continued) -- List courses taken *after* you began full-time teaching. Include undergraduate and graduate courses. (Note: typically one credit-hour corresponds to 10 hours of class time)

	0	1-3	4-9	10-15	16-30	more than 30
a. How many credit-hours of mathematics coursework have you taken?	○	○	○	○	○	○
b. How many credit-hours of physical science coursework (e.g., physics, chemistry, geology) have you taken?	○	○	○	○	○	○
c. How many credit-hours of life science coursework (e.g., biology) have you taken?	○	○	○	○	○	○

	0	1-3	4-9	10-15	16-30	more than 30
d. How many credit-hours of mathematics education coursework have you taken?	○	○	○	○	○	○
e. How many credit-hours of science education coursework have you taken?	○	○	○	○	○	○
f. How many credit-hours of coursework have you taken that focused on the use of questioning, discussion, and inquiry-oriented teaching strategies?	○	○	○	○	○	○
g. How many credit-hours of coursework have you taken that focused on using technology in the classroom?	○	○	○	○	○	○
h. How many credit-hours of coursework have you taken that focused on the use of formative assessment and evaluation of students' prior knowledge to inform instruction?	○	○	○	○	○	○

*7. What subject was your undergraduate major?

*8. Did you complete an undergraduate minor?

○ Yes ○ No

9. If you answered yes to the previous question, what subject was your undergraduate minor?

*10. Part I-C: Professional development

	More than 100	75 to 100	25 to 74	10 to 24	Less than 10
In the past two years, how many hours of professional development have you had?	○	○	○	○	○

11. Describe the professional development experience that has had the greatest impact on your teaching practice. When did this occur? What about that experience impacted your practice?

12. Describe the professional development experience that you found most personally enriching. When did this occur? What made it personally enriching?

*13. Part II: Beliefs and support
Rate the following with how important to student learning you think each is.

	Very Important	Fairly Important	Somewhat Important	Not Important
a. Provide concrete experiences before introducing abstract concepts.	○	○	○	○
b. Develop students' conceptual understanding of physics.	○	○	○	○
c. Develop students' knowledge of science formulas.	○	○	○	○
d. Expose students to a wide variety of topics.	○	○	○	○
e. Take students' prior understanding into account when making instructional decisions.	○	○	○	○
f. Have students work in groups.	○	○	○	○
g. Have students write up lab procedures.	○	○	○	○

	Very Important	Fairly Important	Somewhat Important	Not Important
h. Have students read or hear about an idea before doing an experiment to confirm it	○	○	○	○
i. Make graphs of data on paper by hand.	○	○	○	○
j. Know of the research on student conceptions.	○	○	○	○

*14. How prepared are you for each of the following activities?

	Very well prepared	Fairly well prepared	Somewhat prepared	Not prepared
k. Use informal questioning to assess student understanding.	○	○	○	○
l. Make instructional decisions based on the research on student understanding.	○	○	○	○
m. Help students take responsibility for their own learning.	○	○	○	○
n. Manage a class of students engaged in inquiry.	○	○	○	○
o. Teach about physical science	○	○	○	○

*15. Part III: School climate and administrative support
Rate the following statements.

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
a. I have the support of my colleagues and principal to try out new ideas.	○	○	○	○	○

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
b. Science teachers at this school regularly meet to discuss student work and lesson planning.	○	○	○	○	○
c. Teachers in this school trust each other.	○	○	○	○	○
d. Teachers in this department trust each other.	○	○	○	○	○
e. I am required to follow rules at this school that conflict with my professional judgment about student learning.	○	○	○	○	○
f. My principal is familiar with my instructional practice.	○	○	○	○	○
g. Teachers' curricular and instructional decisions are supported by the administration.	○	○	○	○	○

*16. Part IV: Diagnoser use
Have you visited the Diagnoser Project web site?

○ Yes ○ No

17. If you answered yes to the previous question, what resources did you find most useful? (Check all that apply)

☐ Lesson planning ☐ Elicitation questions ☐ Teacher /Class reports ☐ Facets and Facet Clusters ☐ Suggested lessons

Done >>

Appendix B: Epistemological Beliefs Assessment for Physical Science

EBAPS version 6.1, 02-01-2006

Part 1

DIRECTIONS: *For each of the following items, please read the statement, and indicate (on the scantron answer sheet) the answer that describes how strongly you agree or disagree.*

A: Strongly disagree B: Somewhat disagree C: Neutral D: Somewhat agree E: Strongly agree

1. Tamara just read something in her science textbook that seems to disagree with her own experiences. But to learn science well, Tamara shouldn't think about her own experiences; she should just focus on what the book says.

2. When it comes to understanding physics or chemistry, remembering facts isn't very important.

3. Obviously, computer simulations can predict the behavior of physical objects like comets. But simulations can also help scientists estimate things involving the behavior of *people*, such as how many people will buy new television sets next year.

4. When it comes to science, most students either learn things quickly, or not at all.

5. If someone is having trouble in physics or chemistry class, studying in a better way can make a big difference.

6. When it comes to controversial topics such as which foods cause cancer, there's no way for scientists to evaluate which scientific studies are the best. Everything's up in the air!

7. A teacher once said, "I don't *really* understand something until I teach it." But actually, teaching doesn't help a teacher understand the material better; it just reminds her of how much she already knows.

8. Scientists should spend almost all their time gathering information. Worrying about theories can't really help us understand anything.

9. Someone who doesn't have high natural ability can still learn the material well even in a hard chemistry or physics class.

10. Often, a scientific principle or theory just doesn't make sense. In those cases, you have to accept it and move on, because not everything in science is supposed to make sense.

11. When handing in a physics or chemistry test, you can generally have a sense of how well you did even before talking about it with other students.

A: Strongly disagree B: Somewhat disagree C: Neutral D: Somewhat agree E: Strongly agree

12. When learning science, people can understand the material better if they relate it to their own ideas.

13. If physics and chemistry teachers gave *really clear* lectures, with plenty of real-life examples and sample problems, then most good students could learn those subjects without doing lots of sample questions and practice problems on their own.

14. Understanding science is really important for people who design rockets, but not important for politicians.

15. When solving problems, the key thing is knowing the methods for addressing each particular type of question. Understanding the "big ideas" might be helpful for specially-written problems, but not for most regular problems.

16. Given enough time, almost everybody could learn to think more scientifically, if they really wanted to.

17. To understand chemistry and physics, the formulas (equations) are really the main thing; the other material is mostly to help you decide which equations to use in which situations.

Part 2

<u>DIRECTIONS</u>: *Multiple choice. On the answer sheet, fill in the answer that best fits your view.*

18. If someone is trying to learn physics, is the following a good kind of question to think about?

Two students want to break a rope. Is it better for them to (1) grab opposite ends of the rope and pull (like in tug-of-war), or (2) tie one end of the rope to a wall and both pull on the other end together?

(a) *Yes, definitely*. It's one of the best kinds of questions to study.
(b) *Yes, to some extent*. But other kinds of questions are equally good.
(c) *Yes, a little*. This kind of question is helpful, but other kinds of questions are more helpful.
(d) *Not really*. This kind of question isn't that great for learning the main ideas.
(e) *No, definitely not*. This kind of question isn't helpful at all.

19. Scientists are having trouble predicting and explaining the behavior of thunder storms. This could be because thunder storms behave according to a very complicated or hard-to-apply set of rules. Or, that could be because some thunder storms don't behave consistently according to *any* set of rules, no matter how complicated and complete that set of rules is.

In general, why do scientists sometimes have trouble explaining things? Please read all options before choosing one.

(a) Although things behave in accordance with rules, those rules are often complicated, hard to apply, or not fully known.
(b) Some things just don't behave according to a consistent set of rules.
(c) Usually it's because the rules are complicated, hard to apply, or unknown; but sometimes it's because the thing doesn't follow rules.
(d) About half the time, it's because the rules are complicated, hard to apply, or unknown; and half the time, it's because the thing doesn't follow rules.
(e) Usually it's because the thing doesn't follow rules; but sometimes it's because the rules are complicated, hard to apply, or unknown.

20. In physics and chemistry, how do the most important formulas relate to the most important concepts? Please read all choices before picking one.

(a) The major formulas summarize the main concepts; they're not really separate from the concepts. In addition, those formulas are helpful for solving problems.
(b) The major formulas are kind of "separate" from the main concepts, since concepts are *ideas*, not equations. Formulas are better characterized as problem-solving tools, without much conceptual meaning.
(c) Mostly (a), but a little (b).
(d) About half (a) and half (b).
(e) Mostly (b), but a little (a).

21. To be successful at *most things in life*...

(a) Hard work is much more important than inborn natural ability.
(b) Hard work is a little more important than natural ability.
(c) Natural ability and hard work are equally important.
(d) Natural ability is a little more important than hard work.
(e) Natural ability is much more important than hard work.

22. To be successful at *science*...

(a) Hard work is much more important than inborn natural ability.
(b) Hard work is a little more important than natural ability.
(c) Natural ability and hard work are equally important.
(d) Natural ability is a little more important than hard work.
(e) Natural ability is much more important than hard work.

23. Of the following test formats, which is best for measuring how well students understand the material in physics and chemistry? Please read each choice before picking one.

(a) A large collection of short-answer or multiple choice questions, each of which covers one specific fact or concept.
(b) A small number of longer questions and problems, each of which covers several facts and concepts.
(c) Compromise between (a) and (b), but leaning more towards (a).
(d) Compromise between (a) and (b), favoring both equally.
(e) Compromise between (a) and (b), but leaning more towards (b).

Part 3

DIRECTIONS: *In each of the following items, you will read a short discussion between two students who disagree about some issue. Then you'll indicate whether you agree with one student or the other*

24.

Brandon: A good science textbook should show how the material in one chapter relates to the material in other chapters. It shouldn't treat each topic as a separate "unit," because they're not really separate.

Jamal: But most of the time, each chapter is about a different topic, and those different topics don't always have much to do with each other. The textbook should keep everything separate, instead of blending it all together.

With whom do you agree? Read all the choices before circling one.

(a) I agree almost entirely with Brandon.
(b) Although I agree more with Brandon, I think Jamal makes some good points.
(c) I agree (or disagree) equally with Jamal and Brandon.
(d) Although I agree more with Jamal, I think Brandon makes some good points.
(e) I agree almost entirely with Jamal.

25.

Anna: I just read about Kay Kinoshita, the physicist. She sounds naturally brilliant.

Emily: Maybe she is. But when it comes to being good at science, hard work is more important than "natural ability." I bet Dr. Kinoshita does well because she has worked really hard.

Anna: Well, maybe she did. But let's face it, some people are just smarter at science than other people. Without natural ability, hard work won't get you anywhere in science!

(a) I agree almost entirely with Anna.
(b) Although I agree more with Anna, I think Emily makes some good points.
(c) I agree (or disagree) equally with Anna and Emily.
(d) Although I agree more with Emily, I think Anna makes some good points.
(e) I agree almost entirely with Emily.

26.

Justin: When I'm learning science concepts for a test, I like to put things in my own words, so that they make sense to me.

Dave: But putting things in your own words doesn't help you learn. The textbook was written by people who know science really well. You should learn things the way the textbook presents them.

(a) I agree almost entirely with Justin.
(b) Although I agree more with Justin, I think Dave makes some good points.
(c) I agree (or disagree) equally with Justin and Dave.
(d) Although I agree more with Dave, I think Justin makes some good points.
(e) I agree almost entirely with Dave.

27.

Julia: I like the way science explains things I see in the real world.

Carla: I know that's what we're "supposed" to think, and it's true for many things. But let's face it, the science that explains things we do in lab at school can't really explain earthquakes, for instance. Scientific laws work well in some situations but not in most situations.

Julia: I still think science applies to almost all real-world experiences. If we can't figure out how, it's because the stuff is very complicated, or because we don't know enough science yet.

(a) I agree almost entirely with Julia.
(b) I agree more with Julia, but I think Carla makes some good points.
(c) I agree (or disagree) equally with Carla and Julia.
(d) I agree more with Carla, but I think Julia makes some good points.
(e) I agree almost entirely with Carla.

28.

Leticia: Some scientists think the dinosaurs died out because of volcanic eruptions, and others think they died out because an asteroid hit the Earth. Why can't the scientists agree?

Nisha: Maybe the evidence supports both theories. There's often more than one way to interpret the facts. So we have to figure out what the facts mean.

Leticia: I'm not so sure. In stuff like personal relationships or poetry, things can be ambiguous. But in science, the facts speak for themselves.

(a) I agree almost entirely with Leticia.
(b) I agree more with Leticia, but I think Nisha makes some good points.
(c) I agree (or disagree) equally with Nisha and Leticia.
(d) I agree more with Nisha, but I think Leticia makes some good points.
(e) I agree almost entirely with Nisha.

29.

Jose: In my opinion, science is a little like fashion; something that's "in" one year can be "out" the next. Scientists regularly change their theories back and forth.

Miguel: I have a different opinion. Once experiments have been done and a theory has been made to explain those experiments, the matter is pretty much settled. There's little room for argument.

(a) I agree almost entirely with Jose.
(b) Although I agree more with Jose, I think Miguel makes some good points.
(c) I agree (or disagree) equally with Miguel and Jose.
(d) Although I agree more with Miguel, I think Jose makes some good points.
(e) I agree almost entirely with Miguel.

30.

Jessica and Mia are working on a homework assignment together...

Jessica: O.K., we just got problem #1. I think we should go on to problem #2.

Mia: No, wait. I think we should try to figure out why the thing takes so long to reach the ground.

Jessica: Mia, we know it's the right answer from the back of the book, so what are you worried about? If we didn't understand it, we wouldn't have gotten the right answer.

Mia: No, I think it's possible to get the right answer without really understanding what it means.

(a) I agree almost entirely with Jessica.
(b) I agree more with Jessica, but I think Mia makes some good points.
(c) I agree (or disagree) equally with Mia and Jessica.
(d) I agree more with Mia, but I think Jessica makes some good points.
(e) I agree almost entirely with Mia.

Appendix C: EBAPS Subscales and Scoring

Which EBAPS items belong to which subscales?

Axis 1: Structure of scientific knowledge
2, 8, 10, 15, 17, 19, 20, 23, 24, 28

Axis 2: Nature of knowing and learning
1, 7, 11, 12, 13, 18, 26, 30

Axis 3: Real-life applicability
3, 14, 19, 27

Axis 4: Evolving knowledge
6, 28, 29

Axis 5: Source of ability to learn
5, 9, 16, 22, 25

Overall
All questions on the survey, equally weighted.

NOTE — The following items belong to no axis except for Overall: 4, 21

The following pages list all the EBAPS items sorted by axes and include the scoring scheme for each item.

Axis 1: Structure of scientific knowledge

2. When it comes to understanding physics or chemistry, remembering facts isn't very important.

   A = 0, B = 1.5, C = 2.5, D = 3.5, E = 4

8. Scientists should spend almost all their time gathering information. Worrying about theories can't really help us understand anything.

   A = 4, B = 3, C = 1.5, D = 0.5, E = 0

10. Often, a scientific principle or theory just doesn't make sense. In those cases, you have to accept it and move on, because not everything in science is supposed to make sense.

    A = 4, B = 3, C = 2, D = 1, E = 0

15. When solving problems, the key thing is knowing the methods for addressing each particular type of question. Understanding the "big ideas" might be helpful for specially-written problems, but not for most regular problems.

    A = 4, B = 3, C = 2, D = 1, E = 0

17. To understand chemistry and physics, the formulas (equations) are really the main thing; the other material is mostly to help you decide which equations to use in which situations.

    A = 4, B = 3, C = 1.5, D = 0.5, E = 0

19. Scientists are having trouble predicting and explaining the behavior of thunder storms. This could be because thunder storms behave according to a very complicated or hard-to-apply set of rules. Or, that could be because some thunder storms don't behave consistently according to any set of rules, no matter how complicated and complete that set of rules is.

    In general, why do scientists sometimes have trouble explaining things? Please read all options before choosing one.

    (a) Although things behave in accordance with rules, those rules are often complicated, hard to apply, or not fully known.
    (b) Some things just don't behave according to a consistent set of rules.

(c) Usually it's because the rules are complicated, hard to apply, or unknown; but sometimes it's because the thing doesn't follow rules.
(d) About half the time, it's because the rules are complicated, hard to apply, or unknown; and half the time, it's because the thing doesn't follow rules.
(e) Usually it's because the thing doesn't follow rules; but sometimes it's because the rules are complicated, hard to apply, or unknown.

A = 4, B = 0, C = 3, D = 2, E = 1

20. In physics and chemistry, how do the most important formulas relate to the most important concepts? Please read all choices before picking one.
(a) The major formulas summarize the main concepts; they're not really separate from the concepts. In addition, those formulas are helpful for solving problems.
(b) The major formulas are kind of "separate" from the main concepts, since concepts are ideas, not equations. Formulas are better characterized as problem-solving tools, without much conceptual meaning.
(c) Mostly (a), but a little (b).
(d) About half (a) and half (b).
(e) Mostly (b), but a little (a).

A = 4, B = 0, C = 3, D = 2, E = 1

23. Of the following test formats, which is best for measuring how well students understand the material in physics and chemistry? Please read each choice before picking one.
(a) A large collection of short-answer or multiple choice questions, each of which covers one specific fact or concept.
(b) A small number of longer questions and problems, each of which covers several facts and concepts.
(c) Compromise between (a) and (b), but leaning more towards (a).
(d) Compromise between (a) and (b), favoring both equally.
(e) Compromise between (a) and (b), but leaning more towards (b).

A = 0, B = 4, C = 1, D = 2, E = 3

24.
Brandon: A good science textbook should show how the material in one chapter relates to the material in other chapters. It shouldn't treat each topic as a separate "unit," because they're not really separate.
Jamal: But most of the time, each chapter is about a different topic, and those different topics don't always have much to do with each other. The textbook should keep everything separate, instead of blending it all together.
With whom do you agree? Read all the choices before circling one.
(a) I agree almost entirely with Brandon.
(b) Although I agree more with Brandon, I think Jamal makes some good points.

(c) I agree (or disagree) equally with Jamal and Brandon.
(d) Although I agree more with Jamal, I think Brandon makes some good points.
(e) I agree almost entirely with Jamal.

A = 4, B = 4, C = 2, D = 1, E = 0

28.

Leticia: Some scientists think the dinosaurs died out because of volcanic eruptions, and others think they died out because an asteroid hit the Earth. Why can't the scientists agree?

Nisha: Maybe the evidence supports both theories. There's often more than one way to interpret the facts. So we have to figure out what the facts mean.

Leticia: I'm not so sure. In stuff like personal relationships or poetry, things can be ambiguous. But in science, the facts speak for themselves.

(a) I agree almost entirely with Leticia.
(b) I agree more with Leticia, but I think Nisha makes some good points.
(c) I agree (or disagree) equally with Nisha and Leticia.
(d) I agree more with Nisha, but I think Leticia makes some good points.
(e) I agree almost entirely with Nisha.

A = 0, B = 1, C = 2, D = 3, E = 4

Axis 2: Nature of knowing and learning

1. Tamara just read something in her science textbook that seems to disagree with her own experiences. But to learn science well, Tamara shouldn't think about her own experiences; she should just focus on what the book says.

   A = 4, B = 3, C = 1, D = 0.5, E = 0

7. A teacher once said, "I don't really understand something until I teach it." But actually, teaching doesn't help a teacher understand the material better; it just reminds her of how much she already knows.

   A = 4, B = 3, C = 2, D = 1, E = 0

11. When handing in a physics or chemistry test, you can generally have a sense of how well you did even before talking about it with other students.

    A = 0, B = 1, C = 2, D = 3, E = 4

12. When learning science, people can understand the material better if they relate it to their own ideas.

    A = 0, B = 0.5, C = 1, D = 3, E = 4

13. If physics and chemistry teachers gave really clear lectures, with plenty of real-life examples and sample problems, then most good students could learn those subjects without doing lots of sample questions and practice problems on their own.

    A = 4, B = 3, C = 1, D = 0.5, E = 0

18. If someone is trying to learn physics, is the following a good kind of question to think about?

    "Two students want to break a rope. Is it better for them to (1) grab opposite ends of the rope and pull (like in tug-of-war), or (2) tie one end of the rope to a wall and both pull on the other end together?"

    (a) Yes, definitely. It's one of the best kinds of questions to study.
    (b) Yes, to some extent. But other kinds of questions are equally good.
    (c) Yes, a little. This kind of question is helpful, but other kinds of questions are more helpful.
    (d) Not really. This kind of question isn't that great for learning the main ideas.

(e) No, definitely not. This kind of question isn't helpful at all.

A = 4, B = 3.5, C = 1.5, D = 0.5, E = 0

26.

Justin: When I'm learning science concepts for a test, I like to put things in my own words, so that they make sense to me.

Dave: But putting things in your own words doesn't help you learn. The textbook was written by people who know science really well. You should learn things the way the textbook presents them.

(a) I agree almost entirely with Justin.
(b) Although I agree more with Justin, I think Dave makes some good points.
(c) I agree (or disagree) equally with Justin and Dave.
(d) Although I agree more with Dave, I think Justin makes some good points.
(e) I agree almost entirely with Dave.

A = 4, B = 4, C = 2, D = 1, E = 0

30.

Jessica and Mia are working on a homework assignment together...

Jessica: O.K., we just got problem #1. I think we should go on to problem #2.

Mia: No, wait. I think we should try to figure out why the thing takes so long to reach the ground.

Jessica: Mia, we know it's the right answer from the back of the book, so what are you worried about? If we didn't understand it, we wouldn't have gotten the right answer.

Mia: No, I think it's possible to get the right answer without really understanding what it means.

(a) I agree almost entirely with Jessica.
(b) I agree more with Jessica, but I think Mia makes some good points.
(c) I agree (or disagree) equally with Mia and Jessica.
(d) I agree more with Mia, but I think Jessica makes some good points.
(e) I agree almost entirely with Mia.

A = 0, B = 1, C = 2, D = 3, E = 4

Axis 3: Real-life applicability

3. Obviously, computer simulations can predict the behavior of physical objects like comets. But simulations can also help scientists estimate things involving the behavior of people, such as how many people will buy new television sets next year.

A = 0, B = 1, C = 2, D = 3.5, E = 4

14. Understanding science is really important for people who design rockets, but not important for politicians.

A = 4, B = 3, C = 2, D = 1, E = 0

19. Scientists are having trouble predicting and explaining the behavior of thunder storms. This could be because thunder storms behave according to a very complicated or hard-to-apply set of rules. Or, that could be because some thunder storms don't behave consistently according to any set of rules, no matter how complicated and complete that set of rules is.

    In general, why do scientists sometimes have trouble explaining things? Please read all options before choosing one.

    (a) Although things behave in accordance with rules, those rules are often complicated, hard to apply, or not fully known.
    (b) Some things just don't behave according to a consistent set of rules.
    (c) Usually it's because the rules are complicated, hard to apply, or unknown; but sometimes it's because the thing doesn't follow rules.
    (d) About half the time, it's because the rules are complicated, hard to apply, or unknown; and half the time, it's because the thing doesn't follow rules.
    (e) Usually it's because the thing doesn't follow rules; but sometimes it's because the rules are complicated, hard to apply, or unknown.

A = 4, B = 0, C = 3, D = 2, E = 1

27.

Julia: I like the way science explains things I see in the real world.

Carla: I know that's what we're "supposed" to think, and it's true for many things. But let's face it, the science that explains things we do in lab at school can't really explain earthquakes, for instance. Scientific laws work well in some situations but not in most situations.

Julia: I still think science applies to almost all real-world experiences. If we can't figure out how, it's because the stuff is very complicated, or because we don't know enough science yet.

(a) I agree almost entirely with Julia.
(b) I agree more with Julia, but I think Carla makes some good points.
(c) I agree (or disagree) equally with Carla and Julia.

(d) I agree more with Carla, but I think Julia makes some good points.
(e) I agree almost entirely with Carla.

A = 4, B = 4, C = 2, D = 1, E = 0

Axis 4: Evolving knowledge

6. When it comes to controversial topics such as which foods cause cancer, there's no way for scientists to evaluate which scientific studies are the best. Everything's up in the air!

A = 4, B = 4, C = 2, D = 1, E = 0

28.

Leticia: Some scientists think the dinosaurs died out because of volcanic eruptions, and others think they died out because an asteroid hit the Earth. Why can't the scientists agree?

Nisha: Maybe the evidence supports both theories. There's often more than one way to interpret the facts. So we have to figure out what the facts mean.

Leticia: I'm not so sure. In stuff like personal relationships or poetry, things can be ambiguous. But in science, the facts speak for themselves.

(a) I agree almost entirely with Leticia.
(b) I agree more with Leticia, but I think Nisha makes some good points.
(c) I agree (or disagree) equally with Nisha and Leticia.
(d) I agree more with Nisha, but I think Leticia makes some good points.
(e) I agree almost entirely with Nisha.

A = 0, B = 1, C = 2, D = 3, E = 4

29.

Jose: In my opinion, science is a little like fashion; something that's "in" one year can be "out" the next. Scientists regularly change their theories back and forth.

Miguel: I have a different opinion. Once experiments have been done and a theory has been made to explain those experiments, the matter is pretty much settled. There's little room for argument.

(a) I agree almost entirely with Jose.
(b) Although I agree more with Jose, I think Miguel makes some good points.
(c) I agree (or disagree) equally with Miguel and Jose.
(d) Although I agree more with Miguel, I think Jose makes some good points.
(e) I agree almost entirely with Miguel.

A = 0, B = 2, C = 4, D = 2, E = 0

Axis 5: Source of ability to learn

5. If someone is having trouble in physics or chemistry class, studying in a better way can make a big difference.

A = 0, B = 1, C = 2, D = 3, E = 4

9. Someone who doesn't have high natural ability can still learn the material well even in a hard chemistry or physics class.

A = 0, B = 1, C = 2, D = 3, E = 4

16. Given enough time, almost everybody could learn to think more scientifically, if they really wanted to.

A = 0, B = 1, C = 2, D = 3, E = 4

22. To be successful at science...
    (a) Hard work is much more important than inborn natural ability.
    (b) Hard work is a little more important than natural ability.
    (c) Natural ability and hard work are equally important.
    (d) Natural ability is a little more important than hard work.
    (e) Natural ability is much more important than hard work.

A = 4, B = 3, C = 2, D = 1, E = 0

25.
Anna: I just read about Kay Kinoshita, the physicist. She sounds naturally brilliant.
Emily: Maybe she is. But when it comes to being good at science, hard work is more important than "natural ability." I bet Dr. Kinoshita does well because she has worked really hard.
Anna: Well, maybe she did. But let's face it, some people are just smarter at science than other people. Without natural ability, hard work won't get you anywhere in science!
    (a) I agree almost entirely with Anna.
    (b) Although I agree more with Anna, I think Emily makes some good points.
    (c) I agree (or disagree) equally with Anna and Emily.
    (d) Although I agree more with Emily, I think Anna makes some good points.
    (e) I agree almost entirely with Emily.

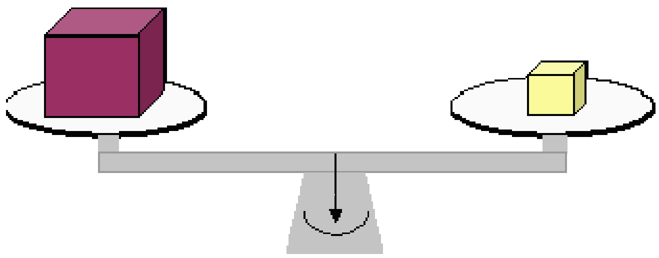
A = 0, B = 1, C = 2, D = 4, E = 4

Appendix D: Paired Questions from Researcher-Designed Pretest and Posttest

Content Area 1: Measuring Mass and Volume

Pretest question:

1. Two blocks shown below are put on either side of an equal-arm balance. The balance remains horizontal.



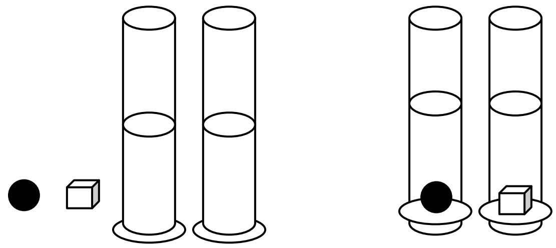
Based on this observation, which of the following are the same for both objects?

Select all that apply.

A. Volume
B. Mass
C. Surface area
D. Density
E. Temperature
F. Number of atoms

Posttest question:

2. Two objects are placed into the graduated cylinders as shown in the pictures. The water level rises to the same height.



Based on this observation, which of the following must be the same for both objects?

Choose all that apply.

A. Density
B. Mass
C. Number of atoms
D. Surface area
E. Temperature
F. Volume
G. None of the above

Content Area 2: Heat and Temperature Basics

Pretest question:

2. As an exercise for a science class, a metal block is put into a refrigerator that is at a temperature of 10^{o}C. A day later the metal block is removed from the refrigerator and immediately placed in a large thermos containing ice water that is at 0^{o}C. Thirty minutes later there is still both water and ice in the thermos bottle and the temperature is still 0^{o}C.

The teacher asks, "Was there heat transfer from the metal block to the ice-water mixture?"

Choose the answer with which you agree the most.

A. There was no heat transfer from the metal block because the block was already cold.
B. There was no heat transfer from the metal block because the ice water stayed at 0^{o}C.
C. There was some heat transfer from the metal block to the ice-water mixture, but too little to be measured by a thermometer.
D. There was some heat transfer from the metal block to the ice-water mixture, but the result was the melting of some of the ice without a change in the temperature.
E. Other. Please describe in the space provided.

Posttest question:

5. As an exercise for a science class, a metal block is left outside on a hot day. The temperature outside is 30^{o}C (86^{o}F). Near the end of the school day, the block is brought inside and is immediately placed in a large thermos containing ice and water that are at 0^{o}C. Thirty minutes later there is still both water and ice in the thermos bottle and the temperature is still 0^{o}C.

The teacher asks, "Was there heat transfer from the metal block to the ice-water mixture?"

Choose the answer with which you agree most.

A. There was no heat transfer from the metal block because the ice-water just made the block colder.

B. There was no heat transfer from the metal block because the ice-water stayed at 0^{o}C.

C. There was some heat transfer from the metal block to the ice-water mixture, but increase in temperature was too little to be recorded by the thermometer.

D. There was some heat transfer from the metal block to the ice-water mixture, but the result was the melting of some of the ice without changing the temperature.

E. This situation is not possible because the temperature of the ice-water would increase a lot if a hot metal block is put in it.

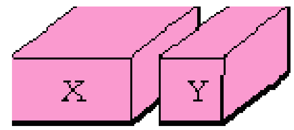
Explain your reasoning in the space below.

Content Area 3: Density

Pretest question:

3. A block of clay is cut into two different size pieces, labeled X and Y.

How does the density of X compare to the density of Y?



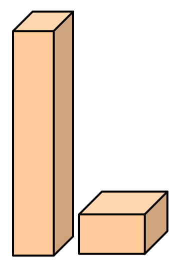
A. The density of X is greater than the density of Y.
B. The density of X is less than the density of Y.
C. The density of X is equal to the density of Y.
D. Not possible to compare without additional information.

In the space below, explain your answer.

Posttest question:

3. A block of aluminum has been cut into two different size pieces, labeled 1 and 2.

How does the density of 1 compare to the density of 2?



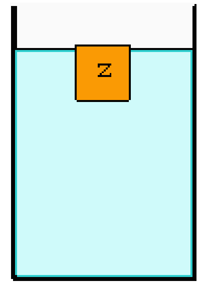
A. The density of 1 is greater than the density of 2.
B. The density of 1 is equal to the density of 2.
C. The density of 1 is less than the density of 2.
D. Not possible to compare without additional information.

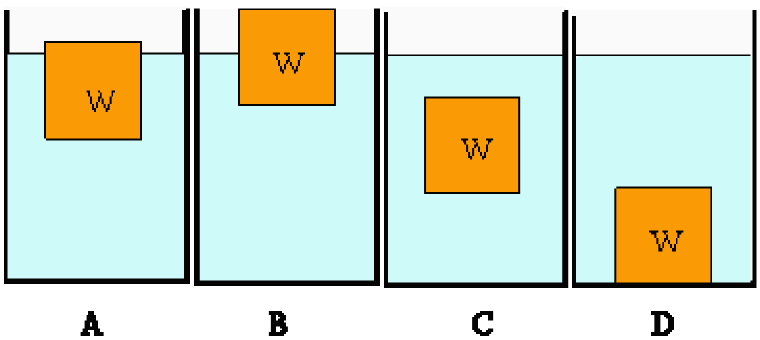
In the space below, explain your answer.

Content Area 4: Sinking and Floating

Pretest question:

5. A plastic block is cut into two pieces, piece W and piece Z. Piece W is larger in size than piece Z. When placed in a cup of water, the smaller piece (Z) floats as shown at right. Which of the following drawings best describes the larger piece (W) in a cup of water?



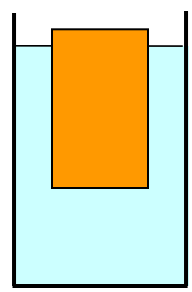


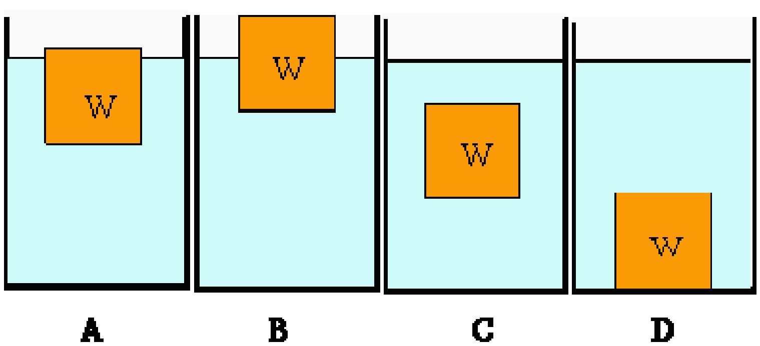
Explain your selection.

Posttest question:

4. A plastic block is cut into two pieces, piece W and piece Z. Piece W is smaller in size than piece Z. When placed in a cup of water, the larger piece (Z) floats as shown in the picture.

Which of the following drawings best describes the smaller piece (W) in a cup of water?



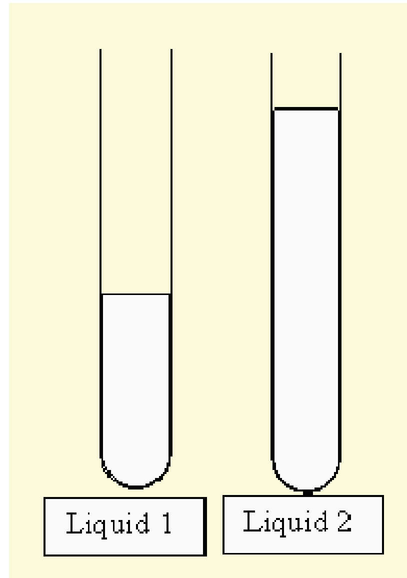


Explain your selection.

Content Area 5: Using Characteristic Properties

Pretest question:

7. A student has two test tubes, each containing a clear liquid. Both samples are at the same temperature. The student wants to determine whether the two samples are the same liquid.



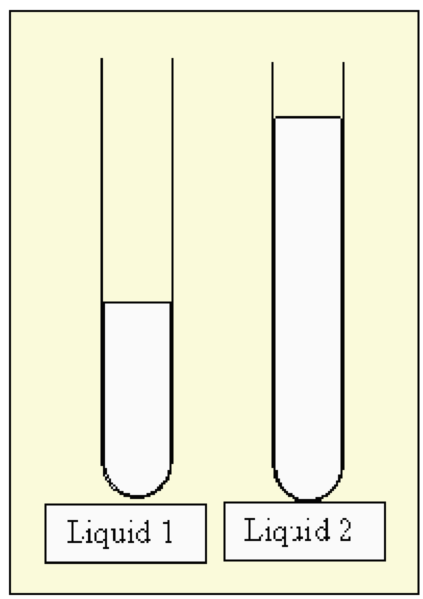
She measures the following properties for each sample. If the two samples are the same liquid, which properties will have the same value for Liquid 1 and Liquid 2?

Select all that apply:

A. Mass
B. Density
C. Volume
D. The maximum mass of salt that can be dissolved completely in each milliliter (mL) of liquid
E. The temperature at which each liquid boils
F. The temperature at which each liquid freezes
G. None of the above

Posttest question:

13. A student has 200 mL of Liquid 1 and 400 mL of Liquid 2. Both samples are clear, colorless, and at the same temperature. The student wants to determine whether the two samples are the same liquid.



She measures the following properties for each sample. If the two samples are the same liquid, which properties will have the same value for Liquid 1 and Liquid 2?

Choose all that apply.

A. The mass of one milliliter (mL) of each liquid
B. The density of each liquid
C. The amount of time it takes to heat each liquid up until it boils using the same burner.
D. The maximum mass of salt that can be dissolved completely in one milliliter (1 mL) of each liquid.
E. The temperature at which each liquid boils.
F. The temperature at which each liquid freezes.
G. None of the above.

Content Area 6: Characteristics of Physical and Chemical Changes

Pretest question:

11. Each of the following describes changing an object in some way. Select all the changes that can be undone (i.e., you can get back the original object).

A. A cup of ice melting
B. Breaking a glass cup into pieces
C. Cooking an egg
D. Burning a piece of paper
E. Mixing Kool-Aid with water

For each of the changes you selected, describe, in the space below, how you would get back the original object.

Posttest question:

10. Each of the following describes changing an object in some way. Choose all the changes that can be undone (i.e., you can get back the original object).

A. A toy house made of Legos is broken into individual Legos.
B. A new shovel gradually gets rusty.
C. Kool-Aid powder is stirred into a jug of water.
D. A few drops of liquid water in a closed water bottle are left in the sun until no liquid water is visible in the water bottle.
E. A teaspoon of coarse black sand is mixed with a teaspoon of fine white sand.
F. Baking soda is mixed with vinegar in a small model of a volcano.

Content Area 7: Periodicity of Elements

Pretest question:

12. Which of the following statements best describes why Nickel (symbol: Ni) and Copper (symbol: Cu) are right next to each other in the periodic table?

A. The density of copper is just slightly greater than the density of nickel.

B. Copper has the next greatest atomic mass after nickel.

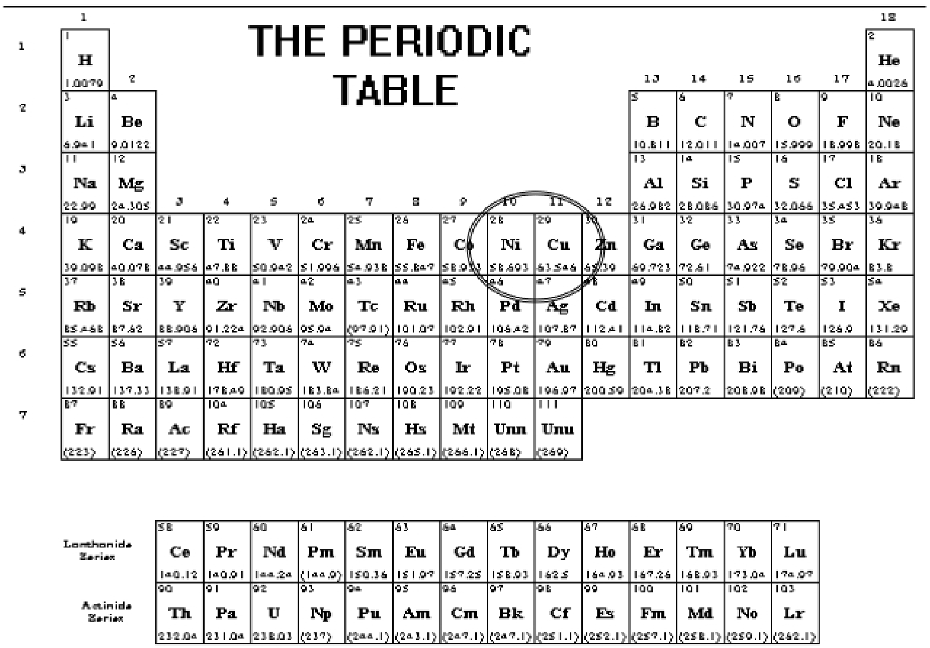
C. A copper atom contains one more proton than a nickel atom.

D. A copper nucleus contains one more neutron than a nickel nucleus.

E. The chemical properties of copper are almost identical to those of nickel.

F. Nickel was the next metal discovered after copper.

Explain your selection.



Posttest question:

11. Which of the following statements **best** describes why Chlorine (symbol: Cl) and Argon (symbol: Ar) are right next to each other in the periodic table?

    A. Chlorine reacts with many of the same elements as Argon.
    B. Argon has the next greatest atomic mass after Chlorine.
    C. An Argon atom contains one more proton than a Chlorine atom.
    D. An Argon atom contains one more neutron than a Chlorine atom.
    E. The odor and color of Chlorine is similar to that of Argon.
    F. Argon was the next element discovered after Chlorine.