AN ABSTRACT OF THE DISSERTATION OF

Matthew F. Price for the degree of Doctor of Philosophy in Science Education presented on July 18, 2008.

Title: Student Views on Science and the Scientific Process: Studying Changes Made in A Redesigned Non-Major Introductory Science Course

Abstract approved:

Lawrence B. Flick

The purpose of this study was to understand how instruction in an introductory non-major astronomy course at Oregon State University affects how non-science majors view science and the scientific process. The study used research-based methods to design a reform-based lecture/lab course and implement it for students. Two top-level questions were asked during the study of the redesigned astronomy course:

1. Do student epistemologies change after instruction using the redesigned lecture/lab curriculum in a non-major introductory astronomy course?
2. Is there any change in content gain when students are instructed using the redesigned non-major introductory astronomy course?

Students were given pre-instruction and post-instruction epistemological measures in order to track changes in their views on science. A pre-instruction diagnostic test was

given to gauge general knowledge, and examination scores were used to assess student content gain.

An initial study of student epistemologies in 2004 indicated a significant decrease in the sophistication of student epistemologies after taking the astronomy course. After our instruction, student epistemologies do not show the decrease that we found initially and in some aspects show a modest increase in sophistication.

We also found no significant decrease in student content gain after instruction was redeveloped to focus on epistemological instruction. In general, as instruction became more epistemologically focused, we found an increase in exam averages over the course of study.

We found a correlation between our instruction and changes in student views about science and the scientific process, but also found content gain to increase.

Student Views on Science and the Scientific Process: Studying Changes Made in A Redesigned Non-Major Introductory Science Course

by

Matthew F. Price

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented July 18, 2008

Commencement June, 2009

Doctor of Philosophy dissertation of Matthew F. Price presented on July 18, 2008.

APPROVED:

____

Major Professor, representing Science Education

____

Chair of the Department of Science and Mathematics Education

____

Dean of the Graduate School

I understand that my dissertation will part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

____

Matthew F. Price, Author

ACKNOWLEDGEMENTS

The list of acknowledgements has to start with my wife and children; sorry for the crazies recently, but it is almost over. Next I would like to thank Larry Flick, Phil Siemens, Shawn Rowe, and Henri Jansen; each one of you has contributed to my knowledge in your own special way. Third would be Zlatko Dimcovic great conversations.

The rest of my thanks go to all of the graduate students that I have interacted with and come to know as friends in both Science and Mathematics Education and Physics; I'll miss the card games guys.

Finally, thanks should go to 3,000 plus students that I have had the pleasure of interacting with at Oregon State University. Their patience and their perseverance is the reason that this document sits on the shelf. Each and every student contributed to this document in some way.

CONTRIBUTION OF AUTHORS

Dr. Lawrence Flick, Dr. Philip J. Siemens, and Dr. Shawn Rowe contributed to the writing of the manuscripts that make up Chapters 2 and 3 of this dissertation. Dr. Siemens contributed to most of the data analysis in Chapter 3. Dr. Flick and Dr. Siemens contributed to the design of the instruction and the implementation of the instruction.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure | Page

LIST OF FIGURES (Continued)

Figure | Page

LIST OF TABLES

DEDICATION

For

Cheryl, ever patient

Sara, the first

Jordan, humor unmatched

Ralph and Ruth

Vera

Granger

Todd and Stephanie

Mark and Reta

Carrie and Roger

And

Mom

Student Views on Science and the Scientific Process: Studying Changes Made in A Redesigned Non-Major Introductory Science Course

CHAPTER ONE

Introduction

In the United States more than 200,000 students yearly take an introductory astronomy course, often the only science course they will have in college (Fraknoi, 2001; Mulvey and Nicholson, 2001). Fraknoi (2001) used electronic submissions from instructors at varying types of institutions – community colleges to large research universities – asking whether they have an undergraduate non-major astronomy course, how many sections of this course they offer, and how many students that enroll each year in this course. Using his results and the data submitted to the American Institute of Physics on the number of astronomy degrees presented each year (Mulvey and Nicholson, 2001), Fraknoi (2001) estimated that the yearly enrollment in non-major introductory astronomy courses is likely greater than 250,000 in the United States. This is a significantly large population of college students for whom this may be their last academic interaction with science. Understanding how introductory astronomy can support science literacy of non-majors will contribute to the research base on building a science-literate citizenry. Instruction designed to help students with astronomy content has been effective (Adams, Prather, and Slater, 2005; Prather *et al.* 2004; Zeilik *et al.*, 1997). Michael Zeilik (1997) developed the conceptual astronomy course at the University of New Mexico where students are instructed using concept maps, and groups work through a series of concept clusters designed to show the students the interconnectedness of the concepts in astronomy. At Arizona State University Adams, Prather, and Slater (2005) developed a set of tutorials for the introductory non-major astronomy lecture course. Research on content and concept measures designed by Zeilik (1997) and Prather *et al.* (2004) show significant pre-post content gains after reformed instruction practices in astronomy.

The research literature on introductory astronomy suggests that a second area of concern affecting non-majors' science literacy is their understanding of the nature of knowledge and knowing in science. Students entering science courses will often have views on how knowledge is gained and transferred to the physical sciences that do not support effective learning. Addressing these views is a critical element of science literacy education (Elby, 2001; Hammer, 1994; Redish, Saul, and Steinberg, 1998; Schommer, 1990).

In the study of students in the science classroom, epistemology is understood as both how they view their own process of gaining knowledge in the classroom (epistemology of learning) and how scientists develop and use scientific knowledge (epistemology of science). As Ryder (1999) notes, epistemology of science consists of understanding many different facets of science. One facet is knowledge of the methods of science – measurement, laws, models, theories, and concepts used by scientists. These topics are often elements of science instruction. Yet they are rarely treated as an integrated process by which scientific knowledge is developed, used and modified. Therefore students in the college classroom have epistemologies of science that do not reflect the underlying epistemology of scientific knowledge (Abd-El-Khalik, 2004). The study by Abd-El-Khalik (2004) indicates that students think that science itself progresses in a manner similar to what we would find in the normal college classroom. In the following discussion I will introduce the epistemology of science. I will also lay out why it is difficult to measure students' epistemology of science. A lengthy discussion on the epistemology of science is beyond the scope of this study.

The epistemology of science is a theory of knowledge about the methods or grounds of science as a way of knowing, or the "values and beliefs inherent to scientific knowledge" (Lederman, 2004). Norman Lederman (2004) outlines an epistemology of science in his discussion of the nature of science. His seven aspects of the epistemology of science are that scientific knowledge is

1. Tentative: Some may think that once scientists accept an idea it is permanent. This view is naïve compared to the accepted view that scientific knowledge can be changed when further study requires this change. In the classroom, the traditional lecture method involves the rapid passing of facts from the teacher to the student, reinforcing the idea that science is base solely on immutable facts.
2. Empirically-based: At least some of our knowledge must be based on measurements of nature. Instruction in a laboratory-based class can emphasize this aspect of the epistemology of science. However, as Abd-El-Khalick (2004) found this may lead to the refusal of any other method for gaining scientific knowledge. Magnusson, Palinscar, and Templin (2004) refer to this view as the methodological model. Knowledge is gained by simply measuring and recording nature.

Students in the classroom may not see the subjective, inferential, imaginative, and creative nature of science.

3. Subjective: Science is subject to the interpretation, based on theory, of the individual or group. Magnusson, Palinscar, and Templin (2004) refer to this as the dialectical model of science. They warn that we should not think of subjectivity in purely relativistic terms, but that the scientific knowledge is based upon the values and beliefs of a community of practitioners who have agreed on what should be studied and interrupted.
4. Involves inference, imagination, and creativity: Science is not limited to dry, orderly, lifeless activities, but sometimes requires the invention of explanations (Lederman, 2004). According to Lederman (2004) this takes a “great” deal of creativity and imagination. The Einsteins’ work on relativity needed many “thought experiments” that could not be done in the laboratory. While students accept a scientist’s creativity in designing experiments, they are less likely to accept creativity in analyzing the data from those experiments (Lederman, 2004).
5. Socially and culturally embedded: A scientist's choice of topic, and of how to interpret the data, does not happen in a vacuum. There are social and cultural influences on that scientist – religious, socioeconomic, political, and other factors. Yet there are also factors from the culture of science. A scientist may wish to study an aspect of

> his field that other scientists in his field have decided is not worth pursuing. There will be limited support from the scientific community for that research. Thus scientific knowledge is negotiated within the community. As discussed with the empirical-based nature of science, Abd-El-Khalick (2004) found that students often ignore this aspect of the nature of science.

The final two aspects of epistemology of science are the distinction between observation and inference, and the distinction between theories and laws (Lederman, 2004). These two differences are the most complex and subtle of the seven. Novice students will not have a belief about either difference. Even students majoring in science may not have sophisticated ideas about these last two aspects of epistemology of science until they have been scientists for many years.

The above framework is complex and sophisticated. As a science major, I personally was not able to articulate my ideas in this framework until I had been in science for many years. Should we expect novice learners to be able to articulate these ideas in the complex framework? Could students hold these views, yet not be able to articulate them? Is there a way to get at students' beliefs about the epistemology of science through an indirect method? It may be that to investigate student epistemologies of science, we must investigate student epistemologies of learning science as well. My study uses surveys of personal epistemologies for learning science to infer student epistemologies of science.

Much has been said about student epistemologies of learning science in the college classroom (Elby, 2001; Hammer, 1994; Redish, Saul, and Steinberg, 1998;

Schommer, 1990). Schommer (1990) found that students with naïve epistemologies will also struggle with comprehension when reading psychology and science texts. While interviewing introductory college physics students, Hammer (1994) found that many students have a naïve epistemology, learning physics in the classroom by memorizing formulas and facts without looking for any connections between these, copying the exact notes of the instructor without adding personal meaning , and attending only to the correct answers for the problems without understanding the underlying concepts. Following a single student through an introductory physics course, Lising and Elby (2005) found that a student who has a naïve epistemology struggles with understanding and retaining content.

Hammer (1989) also found that there are students who have more sophisticated epistemological beliefs about science. During class these students will look for the connection from one topic to the other, try to understand how and why the formulas or facts are connected to the whole of science knowledge, attempt to make their own meaning from the classroom discussions and assigned homework, and try to understand the concepts that underlie the homework problems.

The findings of Hammer (1994) and Lising and Elby (2005) show that student epistemological beliefs are related to classroom learning behaviors. Yet most physics instruction does not foster positive changes in student epistemologies (Hammer, 1989; Redish, Saul, and Steinberg, 1998). Hammer (1989) followed two students, one identified as holding naïve epistemologies, the other holding sophisticated epistemologies. He found that the sophisticated student struggled with learning in the physics course, while the naïve student did better. It is only when the sophisticated student abandoned her sophisticated epistemological stance and adopted the traits of the naïve student that she began to do better in the course. Redish, Saul, and Steinberg (1998) found similar results on student epistemologies across the spectrum of physics courses from traditional lecture/lab courses to well-researched reform curricula. After instruction, students have become less sophisticated in their beliefs about how one learns science. This problem raises interesting issues in the non-major introductory astronomy course.

First, the students who choose a course of this nature are non-science oriented. The major impact that Zeilik *et al.* (1997) and Prather *et al.* (2004) wish to have is on student conceptual understanding of astronomy. Yet as Zeilik *et al.* (1997) write:

> A significant number of these students will become our politicians, teachers, and business leaders. Most either are or will become parents of our next generation. These students and their children will make decisions that will affect the future of our country. Thus the introductory astronomy course represents an important opportunity for undergraduates to develop an understanding of some basic concepts and processes in science, as well as an appreciation of its value.

Instruction that fosters student epistemologies of science will have a long term impact on students who will not have another science class in their college career. Students who show degradation in their epistemologies after a science class may be likely to make less-informed decisions.

The second issue that arises from the problem of losses in student epistemologies is the recruitment and retention of students in the sciences. Anecdotally, if the population of my class is an indicator of the populations of similar classes, there is a small population (around 5%) of freshmen enrolled each fall that have undeclared majors. Negative changes in their scientific epistemologies may lead them away from science. Studies of career goals by Lent, Lopez, and Bieschke (1991) and Luzzo *et al.* (1999) suggest a connection between the ways students view their ability to perform in science and mathematics and their career choices and goals.

In this work I will discuss the design, implementation, and impact of epistemologically based instruction in the introductory non-major astronomy course. I will lay out a framework for instructional design building to the problem that this design process addresses. Following that will be a manuscript that describes the development and implementation of the instruction through each phase of the implementation. Following the manuscript on design will be a manuscript that addresses the effect of the redesigned instruction upon the students in the introductory non-major astronomy course at Oregon State University. Finally I will present a conclusion that connects the two manuscripts together with a discussion on the implications of this research and what steps must be taken next.

Framework

Epistemology

Epistemology as defined by Webster's Dictionary (Merriam-Webster Online, 2008) is "the study or theory of the nature and grounds of knowledge especially with reference to its limits and validity". Research, including Perry's 1970 seminal work in the field of epistemology, suggests that the students entering college will often have varying views – from naïve to sophisticated – on how knowledge is gained and transferred. In the physical sciences Hammer (1994) and Redish, Saul, and Steinberg (1998) show the same continuum of epistemologies. Hammer (1994) in his interviews with six college physics students found the naïve and sophisticated epistemologies defined in table 1. These epistemologies are not binary with all students being in one or the other column, but a continuum of beliefs. The differences in sophisticated and naïve can be clarified by the following examples based on the work of Hammer (1994).

If we were to follow an epistemologically naïve student through the physics course what might we see? First, in the classroom we would see this student diligently writing down everything that the instructor writes on the board. Where the instructor makes a side note, she makes a side note. Where the instructor circles or underlines, she circles or underlines. Later when studying for an exam, the naïve student will take out these notes and memorize them. If we were to ask her why she believes a scientific law, she might respond "The teacher said it was right." During problem solving, say an acceleration problem, the naïve student will sift through all of the formulas for acceleration until she finds one that seems to fit the problem. If that first formula doesn't give the "right answer" she will try another. This guess/equation/check-for-the-right-answer method of problem solving doesn't attempt to discover the underlying concepts of the problem; it only seeks to find the answer. In my astronomy course this student would follow the lab exercise diligently through to the end without once thinking about the concepts that the exercise is intended to illuminate, nor would this student look for the connection between the lab and the lecture. Answers to conceptual questions on the labs would repeat a passage in the book or be based upon what the lab instructor said, not on the findings in the lab.

On the other hand, the epistemologically sophisticated student would act quite differently from the naïve student described above. Notes would be used to help understand the processes involved. Notes might even be annotated with the students' own understanding of the topic or pages from the reading. If compared, the student's notes and the instructor's notes would be similar in content but not structure. A problem would be analyzed for its conceptual underpinnings and what it was asking about before an answer was sought. In my class, the labs would be viewed as another learning tool, and a concept would be understood before moving to the next section. Conceptual answers would be thoughtful, articulating something that the student learned from the findings in the lab or possibly connecting a topic covered in lecture to the findings in the lab.

Hammer (1989) found both these epistemological types in his study. He found that a naïve student (Liza) approached the physics class as I have described. He also found that initially a sophisticated student (Ellen) approached the physics course similarly to what I have described. Hammer (1989) chose Liza and Ellen because they were similar students. Both had reasonable scores on the SATs, and had A's in mathematics. Liza had an A- in her college chemistry course and an A in her high school physics course. Ellen had no prior physics course.

Hammer (1989) interviewed these students during their physics course work. Initially, he found the learning behaviors in Ellen matched those that we as instructors say we would like to see from students – those I described previously. But Liza fully followed the pattern of a naïve physics student as described above. However, as the course progressed Liza had success in the homework and exams, while Ellen struggled. Ellen's struggle to make physics meaningful for her became pronounced when she scored below the average on the second midterm. Ellen managed to achieve a B in the course only after she abandoned her approach to learning physics and adopted an approach similar to Liza's.

Table 1 ***Naïve and sophisticated epistemologies of physics learning as defined by Hammer (1994).***

Belief	Naïve	Sophisticated
Structure of physics knowledge	A collection of isolated facts and pieces	A single coherent system
Content of physics knowledge	Formulas	Concepts that underlie the formulas
Learning Physics	Receiving information from the authority	An active process of constructing one's knowledge

In the previous discussion we saw that Hammer (1994) could identify naïve and sophisticated epistemologies of learning in the physics course. He found that the naïve student managed to be successful in the course and the sophisticated student only found success when she abandoned her approach and began to do "what everyone else does" in the course. It appears that her sophisticated epistemology of learning in the classroom did not align with the coursework, though it aligned with the approach that an instructor would value. Ellen's epistemology impeded her conceptual gain, but in this case the instruction was predicated on the naïve epistemology. Next I will discuss the interaction of epistemology and concept gain.

Naïve epistemological beliefs can impede conceptual gain in the science course (Lising and Elby, 2005) or lead to more difficulty in model-based reasoning (Gobert and Discenna, 1997). Lising and Elby (2005) followed a single student through an introductory physics course using class work videos, written work, and interviews. They found that this student faced a barrier in going from everyday reasoning to formal reasoning about physics content. In interviews the barrier was identified as epistemological. This barrier also impeded understanding of physics content (Lising and Elby, 2005).

Gobert and Discenna (1997), studying a ninth-grade class during instruction in plate tectonics, found a similar interaction with model-based reasoning and epistemology. Students were assessed on their model-based reasoning skills – spatial knowledge, causal/dynamic knowledge, and inferential knowledge – and their epistemological views about scientific models. Gobert and Discenna (1997) found that students with more sophisticated epistemologies had better model-based reasoning skills.

Barbara Hofer, in the introduction to her text written with Paul Pintrich (Hofer & Pintrich, 2002), describes five main models of personal epistemology. These five models consist of the work of Perry; King and Kitchener; Belenky, Clinchy, Goldberger, and Traule; Baxter Magolda; and Schommer. The most important of these five models for my project are the work of Perry (1970) and the work of Schommer (Schommer, 1990, 2002). Perry's framework is important for historical purposes, but is only a small part of the framework of epistemology that informed my project. Moore (2002) describes Perry's 1970 developmental model of college students' experiences. The positions in Perry's model represent a move from what I am calling naïve to sophisticated epistemological beliefs. Moore (2002) takes Perry's initial stages and focuses them into a structured framework. Student epistemologies progress from absolutism to a commitment to relativism. Absolutism is a perspective that centers on the idea of "absolute truths": There is only one perspective and it must be learned; even when other perspectives are recognized, the student will cling to the notion that there is only one right answer and it is "ours"; the answers of the other perspectives are always wrong. In my astronomy course, this would manifest itself in students wanting me to give them the right answer. Frustration would arise if I gave multiple views on a topic and did not tell them which one was absolutely correct. In this stage the teacher is the authority and the student is there to learn "truth".

Commitment to relativism happens when students begin a shift into the "ethical from the conceptual" (Moore, 2002). Students now make commitments that define their identity in the world. This is a highly sophisticated developmental stage for students, one that may take years of intellectual and emotional growth to achieve. This is not a stage that we achieve in a single stand-alone course.

In Perry's framework there is a point between absolutism and commitment with relativism that is important to this study, contextual relativism. Contextual relativism happens when students begin to accept "my own point of view" when making meaning. Students must acknowledge their own point of view when learning, becoming more reliant on reflection about what is known and what is not known. This is the sophistication of students seeing that they gain knowledge by constructing meaning on their own, and acknowledging missing information; the instructor is no longer authority but facilitator. In my astronomy class, I would expect to see students becoming more independent in the lab sections, using the lab instructor as a source of interpretation instead of the source of the answer.

Perry's model is a developmental framework – students move through each of the stages – with one very important aspect that informs instruction. Moore (2002) states that as learners progress through these positions they change their perspective of the teacher. The teacher is no longer an Authority dispensing the "Truth". The teacher is now a facilitator with the expertise to share in interpreting knowledge. Students become active rather than passive learners. A primary goal in physics and astronomy education research is to move the student from passive learner to active learner (Adams, Prather, and Slater, 2005; Crouch and Mazur, 2001; Prather *et al.*, 2004). The progression of stages in Perry's framework does not take into account the contextual nature of epistemologies. In one instance, my astronomy classroom, students may be in the absolutist stage. However, in a different context, say a class in their major, they may be closer to the contextual relativism stage. In my project I use a framework more closely aligned with that of Schommer (1990) and Hofer (2001).

Schommer (1990) and Hofer (2001) use a model of epistemology as based on a system of beliefs. Schommer (1990, 2002) developed her model based on the concept that personal epistemologies are a set of independent beliefs where students may hold sophisticated views in one belief, but naïve views in others. Schommer (1990) studied two groups of college students – university or junior college – using an epistemological questionnaire and their conceptual understanding reading a passage. She found, through factor analysis of the questionnaire, that student beliefs are structured into four independently-held components:

- Innate ability: one is either born good at science or one cannot learn science. In my classroom a student answering the question “How good at science are you” may respond “very poor”. This would indicate that they do not think that they have the innate ability to learn science. My instruction would need to be developed in such a way to help this student understand “ability” as malleable and improved by effort. This would be done through exercises that help the student gain incremental success.
- Simple knowledge: knowledge is made up of discrete and unambiguous facts. A student in a traditional astronomy classroom may carry this view after instruction, since the traditional lecture method is good at conveying facts. Instruction in my classroom geared to change this belief would show the connection. In short, there would be some overarching theme or themes for the entire course that would always connect the presented material with other concepts in astronomy and other parts of science.
- Quick learning: one either learns it quickly or not at all. In an eleven week course, this epistemological belief is hard to address. Even when a slower pace is taken in the course, there is not enough time for students to digest, internalize, and synthesize the material. My instruction should be geared to helping students learn the material at their own pace. We wish to show them what Siemens (2003) described as the tennis or golf approaches to learning.

For some, learning is like playing tennis, fast-paced and quick. For others, learning is like golf. In golf the player thinks about the shot, and then makes it. The golfer approaches the ball for the next shot thinking that about the previous shot and connecting it to the new shot. There is no accepted speed for learning, but we should show the students that it can be done.

- Certain knowledge: knowledge is certain not tentative. This is an important discussion that I can make in my classroom. However, there is a contextual piece to this dimension. It may be fine for a student in my class to think of Newton's laws as certain, but what they need to know is where Newton's laws do not work well. This is the tentative nature of science that I approach. It is not that all knowledge is either certain or not, but that knowledge is testable and can be changed if needed.

Through her study of student conceptual understanding of a reading passage, Schommer (1990) showed that epistemological beliefs have an effect on students' abilities to acquire knowledge. In most college courses there is some amount of reading. Students' understanding of passages in my lab write-ups will affect how they approach the labs. Also, reading of texts and notes will be influenced by epistemological beliefs. This interference will be seen on student exam scores and other assessments. My redesigned instruction take this into account by addressing explicitly how scientists would interpret the meaning of a passage in the text, or rewording a passage in the lab exercise to show the interpretation.

Schommer's framework for epistemology underlies my research project. I add a belief that students can be sophisticated and naïve depending on the context. I agree that students' epistemologies of learning science are going to be a set of coherent beliefs. However, I think that students do not view science as a whole, but believe physics is different from chemistry is different from biology is different from astronomy and so on. Lising and Elby (2005) also showed that epistemologies are context sensitive when their student showed a different epistemology in the classroom *versus* interviews. These context activated epistemologies are similar to those developed by diSessa (1993), Hammer and Elby (2003), and Louca, Hammer and Elby (2004) as explained next.

In physics, diSessa (1993), Hammer and Elby (2003), and Louca *et al.* (2004) consider epistemological resources as composed of finer-grained resources that are activated depending on context. In this case both naïve and sophisticated epistemologies exist in the student, and depending upon the context one or the other may be activated. diSessa's (1993) framework for an epistemology of physics is built upon the premise that students have various phenomenological knowledge structures involving a few parts that are activated when recognized in a physical system or in the system's behavior or hypothesized behavior. In his study of students in a freshmen physics course, students were interviewed while working through a set of specifically designed problems. diSessa (1993) discovered that students will have much less intuition of physical situations than an expert.

Louca *et al.* (2004) made observations of student behaviors and interactions in a third grade classroom during a discussion on why leaves change color. The intervention was designed by the teacher to be epistemologically motivated. They applied their fine-grained resource framework, as well as the developmental framework of Perry (Moore, 2002) and the belief framework of Schommer (1990) to their study in order to determine which framework was more explanatory and predictive. Louca, Hammer and Elby (2004) suggest that their framework is more explanatory and more predictive, arguing that younger students will favor purpose-driven explanations of physical phenomenon over mechanistic explanations.

Hammer and Elby (2003) developed a framework for describing how students may think of knowledge. This framework is intended to guide development of strategies to help students activate their epistemological resources in the introductory classroom. Their discussion revolves around college courses that promote epistemological changes in physics. Hammer and Elby (2003) say that students may think that knowledge can be:

- *Propagated material*: knowledge is something that is passed from one person to another, where the person giving the knowledge does not lose any knowledge, but the person receiving the knowledge gains. A student that has this belief as their primary epistemology might think that knowledge comes from authority, "Because the teacher said".
- *Free creation*: knowledge is invented. This invention is a conglomeration of prior knowledge that the student may have. Students may claim "I made it up", but they are employing other knowledge already available to them in what they deem a novel way.
- *Fabrication*: similar to free creation, fabricated knowledge comes from prior knowledge. In this case the construction of new knowledge from prior

knowledge is a conscious process. The student is aware that they are "putting the parts together".

- *Direct perception*: "How do I know the sky is blue? Because I see it." Knowledge is constructed from direct observations of the world around the student. Students may think that if they cannot see it, it is only inferred and not real.
- *Inherent*: something is known. Knowledge of this sort may be expressed as "I just know it."

Similar to the discussion of Schommer (1990), Hammer and Elby (2003) say that no individual will be in any one of these categories at all times. Individuals may display different beliefs about knowledge in different contexts or even when shown different content in the same context. The knowledge of the student is situated and constructed in the context.

In the previous discussion, I showed various epistemological frameworks that research has suggested. The work of Perry (Moore, 2002) shows a developmental framework. The work of Schommer (1990), the primary framework I use, shows epistemologies as independent beliefs. diSessa (1993), Hammer and Elby (2003), and Louca *et al.* (2004) also think that epistemologies are independent, but add that they are very fine-grained structures that are activated depending on context. The framework that guided my research consists of the coherent belief system of Schommer (1990), but added the context-sensitive nature of those beliefs similar to diSessa (1993), Hammer and Elby (2003), and Louca *et al.* (2004). It is in this framework that my instruction was developed. However, my instruction also used the lessons of physics education research and astronomy education research. The next sections deal with these findings separately.

Physics Education and Astronomy Education and Epistemology

In this section I will highlight the work of physics education research (PER) and astronomy education research (AER). The literature on PER is extensive, and a complete review is beyond the scope of this study. A comprehensive review of PER, including instruction and curricula, can be found in McDermott and Redish (1999). My primary goal in this section is to highlight the current PER discussion about instruction and curricula reform beginning with the work of Hake (1998). I will continue with a discussion of one instruction method that I used in my design and an example of what PER considers a successful curriculum. In both of these examples some form of scaffolded instruction has been incorporated into the method. With this in mind I will discuss scaffolded instruction. I will then highlight what PER has found about epistemologies of physics students. This will lead into a discussion of AER.

Early physics education research has identified students' difficulties understanding a broad range of physics concepts, through the development of concept inventories and tests. The most used inventory is the Force Concept Inventory (FCI) developed by Hestenes, Wells, and Swackhammer (1992). The FCI is intended to measure student understanding of Newtonian mechanics as well as to measure the levels of misconceptions that students may have in mechanics. Research using the Force Concept Inventory shows that traditional lectures in introductory science courses are relatively ineffective at improving student content knowledge for the majority of physics students (Hake, 1998). Hake (1998) used data from 6,000 students completing the Force Concept Inventory to calculate the normalized gain of students pre- to post-instruction responses. A normalized gain score is (Hake, 1998):

Gain = (post-pre)/(100%-pre)

When Hake (1998) looked at the normalized gain of students he found that there was a difference between gains posted by students in traditional physics instruction – didactic lecture and prescribed labs – and those in instruction with some form of reform teaching. Hake (1998) found that students in the traditional course showed a normalized gain of 0.2, which indicates a positive change on the FCI pre to post. This means that for some, the traditional physics course leads to conceptual gains. However, Hake's normalized gain for a classroom with some reform instruction taking place was 0.48, more than two standard deviations greater than the traditional course. His conclusion was that the traditional method was ineffective at helping a large portion of students to learn the concepts.

In response to Hake (1998) different instruction and curricula have been developed to target what is seen as the underlying reason that students still have conceptual difficulties – traditional forms of instruction are ineffective at helping students come to a deep understanding of the material (Crouch and Mazur, 2001; Laws, 1996; Manogue, Siemens, Tate, Browne, Niess, and Wolfer, 2001). As mentioned in the introduction to this section, many of the methods developed in PER use a form of scaffolded instruction. Often these methods show large increases in normalized gain scores on the Force Concept Inventory (Crouch and Mazur, 2001). A widely adopted instruction technique in physics is Peer Instruction (Mazur 1997). Some of the more well known curricula are: Tutorials in Physics (McDermott 1998);

Workshop Physics (Laws 1996); and Paradigms Physics (Manogue *et al.* 2001). The two examples of instruction and curriculum I will highlight will be Peer Instruction (Mazur 1997) and Workshop Physics (Laws 1996). However, to show how these incorporate scaffolded instruction, I will first discuss scaffolding in general.

The concept of scaffolded instruction is the process by which a student is able to solve a problem, carry out a task, or achieve a cognitive goal that would otherwise be beyond the skills of the student (Palinscar 1986). The term scaffolding – coined by Wood, Bruner, and Ross (1976) – is used as a metaphor. In construction, scaffolding is built around the building and slowly removed as the building becomes complete and able to stand up on its own. Similarly in scaffolded instruction the teacher provides extensive initial support for the student when learning a new topic or task, and slowly removes the support as the student begins to internalize the process cognitively and become independent (Meyer and Turner, 2002). One of the basic premises behind scaffolded instruction is that there is a difference between what the student can accomplish alone compared to what the student can accomplish with guidance from the instructor – the Zone of Proximal Development (Vygotsky, 1978). Palinscar (1996) describes Vygotsky's Zone of Proximal Development as the difference in the actual cognitive level and the potential cognitive level. Palinscar (1986) states that there is one important question in scaffolded instruction: how can teachers aid learners in going from level to level until they can independently apply the taught skill?

An essential ingredient of successful scaffolded instruction, as described by Palinscar (1986), is a complete examination of the objective to be learned and the development of knowledge to meet that objective. The instructor can use this examination of the subject to devise instruction to highlight certain features while maintaining student interest in the subject (Wood *et al.*, 1976; Palinscar, 1986). During scaffolded instruction, there is significant interaction between teacher and student. Considerable questioning, feedback, and evaluation by both the teacher and the student are done. By its nature, scaffolding increases the interaction between teacher and student creating a social learning environment. This creates an apprenticeship between student and teacher. As discussed previously; Perry's developmental model suggests the teacher becomes a facilitator of learning instead of the authority dispensing "Truth'. This argument is born out in the work of Wood *et al.* (1976) and Palinscar (1986).

In PER, scaffolded instruction finds its way into new instruction and curriculum in several ways. In Peer Instruction (Mazur, 1997), described below, the questioning, feedback, and evaluation by both teacher and student take place in a question and answer cycle. The Paradigms Physics curriculum (Manogue *et al.* 2001) students interact in small groups with the instructor and a teaching assistant in the classroom. Students are taken through a series of tasks that guide them to a deeper understanding of the topic of discussion. Both of the methods here also ask students to think about what they are learning and reflect on those thoughts. This is the last aspect of scaffolded instruction that I will discuss.

The final important aspect of scaffolded instruction is metacognition by the student. Metacognition is the process by which the student becomes more aware and responsible for her cognition and thinking (Pintrich 2002). As Pintrich (2002) states: in practical terms, we wish the student to become more aware of the processes that they undertake when trying to learn new material, and what they understand of the material. Most importantly as facilitators, we want the student to become more aware of what parts they don't understand and where to go for help in understanding those parts. The teacher becomes a source for interpretation and help, not the authority dispensing truths. In the classroom; we would like students to reflect on the topic before, during, and after instruction. White and Frederickson (1998) in their study of inquiry had specific reflection assessment where students evaluate their own as well as their peers' research in their inquiry classroom. In this case, reflective assessment is the aspect of metacognition in White and Frederickson's (1998) scaffolded instruction.

In the previous discussion I outlined scaffolded instruction. In physics education, one instructional method and one curriculum that use a form of scaffolded instruction are Peer Instruction (Crouch and Mazur 1998) and *Workshop Physics* (Laws 1996) respectively. Both of these methods are considered to be successful in fostering a deeper understanding of physics in the students. Both also use reflection to foster student metacognition. I will discuss each separately below, Peer Instruction as a model of instruction that I incorporated into my design, and *Workshop Physics* because it will be discussed again as it relates to epistemological studies in PER.

Peer Instruction (PI) has significantly improved both conceptual and computational achievement in students taking calculus-based and algebra-based physics. Developed by Eric Mazur, professor of physics, engineering, and applied sciences at Harvard University, PI has a ten-year history of demonstrated success in courses taught at research universities as well as 2-year colleges. The technique was designed to improve active learning and teaching in large lecture courses for either majors or non-majors. Study data are taken from pre- and post-tests composed of concept questions designed to parallel the course instruction. Assessments taken at Harvard University in calculus-based physics between Fall 1990 and Fall 1997 show a four-fold increase normalized gain scores (Crouch & Mazur 2001). Results using the more widely studied “Force Concept Inventory” show a three-fold increase (Hake, 1998; Hestenes, Wells, & Swackhammer, 1992).

As an example of scaffolded instruction, PI includes lectures designed around short presentations highlighting specifics of the subject, followed by a conceptual question for which students take 2 to 3 minutes to formulate individual answers. The instructor evaluates these answers to determine whether the majority of the group has understood the concept. If there is significant misunderstanding, small groups of students (PI) enter into a structured 2 to 4 minute discussion in an attempt to reach a consensus on the answer and convince others of the soundness of their reasoning. During this time the instructor moves about the room listening to the discussions and urging students to either pursue their line of thinking or gently guiding students down a new path toward the desired concept. The instructor calls the class together, polls groups for their answers and underlying reasoning, then provides a summary explanation of the answer and moves on. Less is covered in lecture but it is covered more effectively. The final component of implementing a PI class is to have students complete a reading assignment before lecture that is closely aligned with the lecture content. Students are given both incentive and guidelines for thinking about the reading content. Students receive credit (5% of course grade) for completing a three-question, Web-based assignment over the reading. Two of the questions probe difficult aspects of reading and the third asks for what they found difficult or interesting. The instructor uses these responses to prepare for the lecture saving preparation time and focusing content on concepts the students found problematic (Crouch & Mazur 2001). This is a major work of instructional scaffolding in the physics education research community.

Workshop Physics (Laws 1996) is a curriculum that uses all laboratory-based implementation in the classroom. Instead of lecture, students meet for 2 to 3 hour sessions where they engage in a series of hands-on demonstrations, data-taking labs, problem-solving sessions, model derivations, and other activities. After their time in *Workshop* Physics, Laws (1989) claims that students reduced their errors on the Mechanics Baseline Test – predecessor of the Force Concept Inventory – from 65% pre-instruction to 12% post-instruction. Scaffolded instruction primarily takes place in the problem solving sessions and the labs. In the problem solving sessions, students are introduced to progressively harder problems on a single topic. At some point the instructor must step in and help. With each progressive problem the instructor decreases the level at which help is given. In the labs a similar situation occurs. Students are given tasks to complete. The lab instructors observe and intercede when the students have reached a point that they are not progressing as a group. Laws (1996) uses students that have previously been in the course as lab instructors. Students currently in the course have peers – possibly in the same major – to assist them creating an apprenticeship feel to the lab classroom.

The above description of PI and *Workshop Physics* shows concrete examples of the forms that scaffolded instruction can take. However, the basics of scaffolded instruction can be identified. Both use significant and structured interactions between learners and teachers. In both cases the peers often act as teachers. In PI the peers are in the same course; in *Workshop Physics* the peers are students that had prior success in the course. Also both use methods of reflection to facilitate student metacognition. PI has the Web-based writing assignments. *Workshop Physics* incorporates reflection into the lab discussion. Both have seen significant gains on instruments measuring student concept knowledge. However, there is another question pertinent to our framework: in what ways have these methods changed students' personal epistemologies about learning science? This is discussed next.

While there have been a large number of research studies in physics education, the few geared to find out how students view science and the scientific process in physics, have mostly concentrated on the introductory physics courses for science and engineering majors (Redish, Saul, and Steinberg 1998, Elby, 2000, Mcdermott and Redish, 1999). Redish, Saul, and Steinberg (1998) used the Maryland Physics Expectation Survey (MPEX) to study student epistemologies and expectations in introductory physics courses in 6 different collegiate settings using various instruction techniques – from traditional lectures with no reform instruction to the well researched *Workshop Physics* program developed by Laws (1996). The MPEX is a Likert-type instrument based on six subscales – independence, coherence, concepts, reality link, math link, and effort – that probe students' beliefs about the physics classroom. The MPEX is scored on those student answers that most match (favorable) the answers of experts on the same questions. Using a pre-post format for their study, Redish, Saul, and Steinberg (1998) found that the percentage of favorable student responses to the

MPEX decreases after instruction, especially in the reality link and effort subscales. The decrease in the reality link subscale indicates that the students felt less connection to reality after the course compared to before the course. The effort decrease indicator that the students felt they would need to work hard and that they did not do this. This finding is consistent across all of the institutions regardless of instruction methods. Even a curriculum such as *Workshop Physics*, which has shown significant gains in content knowledge, shows a net degradation in student epistemologies and expectations. However, the decrease was less than any of the other classrooms that were in the MPEX study. It appears that even well researched curricula that show large student gains in physics content knowledge do not help students to gain a sophisticated epistemology toward physics. No similar studies have been made concerning PI.

Building on the work of Redish, Saul, and Steinberg (1998) physics education has started to study instruction intended to foster epistemological changes. Lising and Elby (2004) state that epistemologically based instruction can promote two goals: a) attending to student views can help to improve instruction by explaining variations seen in the outcomes (as measured by the FCI) of research-based curricula – especially variations seen from course to course and institution to institution; and b) fostering student epistemology could help the students beyond the introductory physics or astronomy courses (diSessa, 1993; Elby, 2001; Hammer, 1994; Hammer and Elby, 2003; Lising and Elby, 2005; Louca Hammer and Elby, 2006; Redish *et al.*, 1998; Zeilik, 1997). As an example Elby (2001) developed a physics course meant to foster positive gains in scientific epistemology for high school students. Highlighted by the use of non-traditional homework and radically reduced content coverage (an aspect of PI as well), his instruction was introduced to high school students in California and Virginia. The California group was from a middle-class high school whereas the Virginia group was from a talented magnet school group. Using the MPEX Elby (2001) showed an increase in all of the subscales defined by Redish *et al.* (1998) except for the effort subscale. There seems to be a disconnect with what the students see as the amount of effort that they need to put into the course and the amount of effort that they feel they really put into the course. Elby's results indicate that instruction designed to foster epistemology, will. There is also an indication in the work of Hammer and Elby (2003) of students' gaining a deeper level of knowledge of the physics concepts.

The previous section highlighted the efforts of PER in recent years. Beginning with Hake (1998), we see that there is a need to transform undergraduate physics instruction. I highlighted PI and *Workshop Physics* as two models that do just that. Both are successful in conceptual gain, but in the case of *Workshop Physics*, Redish, Saul, and Steinberg (1998) showed regression in students' epistemologies. No similar work studying PI has been done. I connected PI and *Workshop Physics* to scaffolded instruction and highlighted where scaffolding takes place in each. The next section begins the discussion of AER and connects it to the discussion of PER.

In this section I will highlight the two astronomy instruction methods that I used in my course design. The method of Zeilik *et al.* (1997) and the method of Prather *et al*. (2004) are similar in some ways but also have differences that distinguish them. I will highlight those differences.

Many professional astronomers, influenced by physics education research, have taken up the study of astronomy education within the non-major introductory course (Adams, Prather, and Slater, 2005; Green 2003; Prather *et al.*, 2004; and Zeilik, 1997). Education research and instructional research in the astronomy community have motivations similar to those of the PER community. Traditional lectures are ineffective at improving content knowledge for a majority of students. Students are expected to synthesize and understand basic and complex science concepts from simple instruction using only lecture, labs, and homework. Also, introductory science courses often cover an overwhelming amount of material (Adams, Prather, and Slater 2005; Prather *et al.*, 2004; Shipman, 2004; Zeilik *et al.*, 1997). The two major instructional models developed to address these problems are conceptual astronomy (Zeilik *et al.*, 1997) and tutorials in astronomy (Prather *et al.*, 2004). Also, at Harvard, Paul Green (2002) has adapted PI for the astronomy course. I will discuss each separately next.

Conceptual astronomy instruction developed by Zeilik *et al.* (1997) at the University of New Mexico specifically targets the large lecture course with no laboratory section. Their approach is based on conceptual understanding of astronomy content, with some attention to cognitive studies in physics and psychology. Zeilik *et al.* (1997) developed instructional techniques that center around concept maps, small group tutorials, and a different type of assessment based on conceptual gains as seen in pre-post diagnostic tests. The highlight of this instructional method is the development of concept clusters. The concept cluster is a group of concepts that share a common theme in physics or astronomy. An example of a concept cluster is “Scientific

Models". The concept may appear in discussions about the Solar System or stars, but Zeilik *et al.* (1997) use the concept cluster as a connecting piece to each of these separate phenomena. Assessment results show an improvement in content and concept pre- to post-instruction.

Adams, Prather, and Slater (2005) and Prather *et al.* (2004) working at the University of Arizona developed a set of lecture tutorials for use in a large lecture non-laboratory introductory astronomy course. The lecture tutorials are designed to actively engage students in the topic of the day with questions posed in a series of increasingly difficult steps akin to scaffolded instruction. A tutorial will take the form of questions or conversations written in everyday language. Students are guided through individual steps taking them from the simple to the complex.

In his study of the tutorials, Ed Prather at the University of Arizona (2004) showed that in a lecture-only astronomy class, significant conceptual gains can be made by introducing lecture tutorials into the standard lecture classroom. The tutorials place more emphasis on the learning of content and concepts, but less on how we construct knowledge from this content or these concepts, or how they become part of scientific knowledge. That is to say that Prather *et al*. (2004) set out to develop conceptual understanding in the students and they were successful. In my instruction, tutorials were used to help facilitate discussion. I viewed lecture tutorials as advanced organizers (detailed in a subsequent manuscript) for my labs, and so I integrated them into the pre-labs.

Paul Green (2002) has adapted PI at Harvard University using the same methodology as Mazur (1997). No formal research has taken place into PI in the astronomy course, but Green (2002) argues that if the implementation of PI in astronomy is similar to PI in physics, the conceptual gains should also be similar. This assumes that the students have the same level of ability entering into the astronomy course as the students entering the physics courses at Harvard.

All three of the previously discussed instruction methods contain few findings that show changes in student understanding of the nature of science. Zeilik *et al.* (1997) developed a Views of Astronomy survey to give to students. They noted that student views about astronomy changed over the course of the term, with students' overall views on astronomy improved after the course. Prather *et al.* (2004) discuss the work of Redish *et al.* (1998). Green (2002) mentions that he wants to change how students view the nature of science. However, like physics education, more attention is paid to conceptual and content change than to changing epistemologies of science.

In the previous discussion of my framework, I have discussed how we can look at epistemology. I have framed the argument for measuring students' epistemologies of science through their epistemologies of learning science. I have highlighted the work of PER and AER in the development of instruction and curricula that have led to significant increases in the conceptual gains of students in physics and astronomy. I have also discussed the decrease in student epistemologies as quantified by Redish *et al.* (1998). Finally I have highlighted the need for epistemology-based instruction in the non-major science course. All of this leads to the problem that defines the scope of my research. I will discuss this next.

The Problem

Most physics and astronomy research into instructional development has been motivated by the need to increase students' knowledge of content. In many cases students are given instruction geared to help their understanding of the content, without special concern for their understanding of the nature of science and the scientific endeavor. With Elby (2001) and Lising and Elby (2005) there have been discussions about instruction designed to foster positive epistemological change in physics students. There has not been a similar discussion in astronomy education research until now. Given that the introductory non-major astronomy course may be the only science course that these students take, it is important that the learning experience for students succeeds in two areas: a) students exiting should have a better understanding of the content and concepts in astronomy; b) students exiting should have a significantly more sophisticated personal and scientific epistemology compared to when they entered the course. Research using content and concept measures designed by Zeilik (1997) and Prather *et al*. (2004) show significant pre-post content gains after reformed instruction practices in astronomy. The next logical step in this field of research is to apply this research to the design and implementation of reformed teaching in introductory astronomy.

We have seen that students entering the non-major introductory science will often have what we would define as naïve views on how knowledge is gained and transferred in the physical sciences (Elby, 2001; Hammer, 1994; Redish, Saul, and Steinberg, 1998; Schommer, 1990), and we have seen that students with more sophisticated epistemologies learn the content and concepts at a deeper level. As instructors we should take steps to develop instruction intended to foster epistemological gains in our students, especially in our non-major introductory science courses. In the physics department and the science and mathematics education department at Oregon State University this instructional modeling is taking place.

The instructional design discussed here is a joint effort between the Department of Physics and the Department of Science and Mathematics Education at Oregon State University. Contributing members to the design process were Philip J. Siemens (Physics) and Lawrence B. Flick (Science and Mathematics Education).

We have developed a method of instruction for our non-major introductory astronomy course based on the inquiry cycle developed by White and Frederickson (1998). The ThinkerTools curriculum developed by White and Frederickson (1998) is based on scaffolded inquiry, reflection and generalization. Their general theory that we used as an approach states that understanding is developed by successive refinement, and instruction proceeds using simplification followed by the limited introduction of complexity. The instructional structure supports students' development not only of cognitive skills for scientific argumentation, but also of insight into the conduct of science through inquiry. The inquiry-based model of instruction follows the theme that authentic science tasks can help students become better aware of the processes that occur in the science community. At the same time, our goal is to help students learn astronomy content on a deeper level through scaffolded instruction (explained previously) not just in an individual task, but for the entire term.

In order to help students gain an appreciation for the nature of science and scientific knowledge, we asked the following questions during each phase of the design of the project:

1. Do student epistemologies change after instruction using the redesigned lecture/lab course in our non-major introductory astronomy course?
2. Is there any change in content gain when students are instructed using our redesigned non-major introductory astronomy course?

In this chapter, I have developed a framework that has defined my research problem and questions. The questions above guided the following manuscripts. The first manuscript will make use of the discussion on epistemology in explaining the framework for the redesigned instruction. Furthermore, the first manuscript will make more use of the instructional methods discussed here to develop a framework for instruction that overarched the entire design process. I will discuss in detail that design process. The second manuscript will make extensive use of the epistemological framework to discuss the implications of the findings of the empirical data collected during this project. The final chapter of this document will connect the two manuscripts in a theme that will help us discuss the future of this project.

CHAPTER TWO

Design and Implementation of an Inquiry-Based Non-Major Introductory Astronomy Course

Matthew F. Price

To be submitted to The Journal of Research in Science Teaching

Abstract

Research on the changes in student scientific epistemologies in the introductory physics and astronomy classroom has indicated that even in the more successful reform physics instruction and curricula, students do not change their views about the structure of science (Redish, Saul, and Steinberg, 1998). However, Elby (2001) showed that instruction in a physics course intended to foster epistemological changes will lead to students thinking about physics in a manner similar to professionals while maintaining content acquisition. Our study follows successful instruction models from physics and astronomy (Mazur, 1997; Prather *et al.*, 2004; White and Fredericksen, 1998, Zeilik *et al.*, 1997) to develop an implement an introductory astronomy course for non-majors at Oregon State University. Guiding our development was the following question: can an epistemologically focused non-major course foster changes in students' epistemologies? We implemented the redesigned instruction in three phases over the 2006-2007 academic years. We found that the application of inquiry-based instruction toward student epistemologies succeeded in helping students change their views on learning science in the classroom.

Introduction

If one wishes students to change their views about the nature of knowing and learning in science, what is the best way to promote that? The short response is, design a class for that outcome. Is it necessary then to design a class on the nature of science, or can we adequately advance this topic in the context of a science class? At Oregon

State University, we began a design project to modify our existing non-major introductory astronomy course to foster student epistemologies of science.

Epistemology is the science or theory that pertains to the process of gaining and incorporating knowledge in society. For us scientific epistemology specifically refers to the process of gaining and incorporating knowledge in science. One can distinguish two types of epistemology of science. First, there is the epistemology of science that is usually referred to as Nature of Science (Lederman, 2004). This epistemology has to do with the process and products of science. In this epistemological framework, science is:

- tentative
- subjective
- empirically-based

and science involves:

- inference, intuition, and creativity
- society and community
- knowledge of the difference between observation and inference
- knowledge of the difference between theory and law.

This is a very subtle and complex framework. In the traditionally presented astronomy classroom, I may be able to cursorily go over each aspect of this framework. However, the students will leave either confused or less likely to believe items from this framework.

The second form of epistemology of science is epistemology of learning science. Students have coherent and specific beliefs about how they learn science in the classroom. This is a personal epistemology. Schommer (1990) and Hammer (1994) have shown that students hold these beliefs. Redish, Saul, and Steinberg (1998) and Elby (2001) have developed instruments to measure the various student epistemologies of learning. In all cases they find that students think of learning science involving:

- Innate Ability: You either are good at learning science or you are not.
- Simple Knowledge: Science is a bunch of unconnected facts you memorize in the class, not interconnected ideas.
- Quick Learning: You either learn science quickly or not at all;
- Certain Knowledge: This is connected to the tentative nature of science. Students think that once we accept an idea it is proven. They do not understand that ideas are subjected to testing and retesting regularly.

This framework can and should be discussed thoroughly in every science classroom. Our goal will be to develop instruction that pries the epistemology of students from the absolutes of this framework.

Reforming a course in the manner described above is best done through a design experiment. Design experiments are conducted in an iterative manner. As instructional reforms are introduced, they are tested and revised as needed. The main purpose of design experiments is to develop or test theories. As Edelson (2002) states the conventional role of design as a strategy for testing theories proceeds through a general sequence of:

1. development of a theory
2. derivation of principles from the theory

3. Translation of the principles into concrete designs
4. Assessment of the designs to test whether they work as anticipated

Design studies are typically where education innovations are tested. Design experiments have a prospective piece, where innovations are designed using prior research and theory. Designs are also reflective, where innovations are tested against prior expectations and theory (Cobb *et al.*, 2003).

This project has entailed the development of an instructional model based upon prior work in the areas of epistemology and education research. In this manuscript, I will describe the steps that were taken to develop each individual innovation that was implemented during the course of this project. Prior to that I will describe the instructional model that was used to design this course.

Design

Introduction

The overall design of this course uses scaffolded instruction to foster changes in student epistemologies of science through changing their epistemologies of how they learn science. The instructional changes were made over the course of the 2006-2007 academic year. At Oregon State University the academic year is three 11 week terms with 10 weeks of class meetings and one final-exam week. The astronomy course that I worked on is a stand-alone 11 week course, so each term we had a new experimental group. We introduced the instructional changes in three phases. The first phase was in the fall term of 2006 (FALL 2006) where we introduced only redesigned labs. The second phase introduced more varied instruction into the lecture class and maintained the labs as they were. The second phase was implemented in the winter term of 2007 (WINTER 2007). The final phase, during spring term of 2007 (SPRING 2007), introduced a pre-lab component that structured the discussion for each week's meetings. Table 2 shows a timeline of the instructional changes made. The rest of this section is devoted to discussing the setting and expanding on the instructional styles mentioned in this paragraph.

Table 2 ***Timeline of instructional changes*****.**

Term	Instructional Changes	Instructional content
FALL term 2006 (FALL 2006)	Introduced redesigned labs	Lecture instruction using only Peer Instruction for entire term. Lab instruction using scaffolded instruction based on the inquiry cycle of White and Fredericksen (1998)
Winter Term 2007 (WINTER 2007)	Maintained labs from FALL 2006, introduced modified instructional practices in lecture	Lecture instruction using Peer Instruction, concept maps, tutorials, advanced organizers, demonstrations. Maintained lecture pace regardless of lab.
Spring Term 2007 (SPRING 2007)	Introduced pre-lab that coupled instruction from WINTER 2007 and labs from FALL 2006.	Pre-labs designed as part of inquiry cycle to prepare for lab. Pre-labs used as an instructional tool in lecture class along with other lecture methods.

Setting

The non-major introductory astronomy course is a stand-alone eleven week course offered every term of the academic year. There are ten weeks of class meetings followed by one final exam during week eleven. The course enrolls 260 students per term. The lecture portion of the course meets three times per week for fifty minutes each session, in which all students gather in a large lecture hall with one instructor (me). Smaller groups consisting of 26 students meet with a course teaching assistant, acting as laboratory instructor, once per week for two hours in a computer-based laboratory.

The lecture hall is typical of large enrollment courses. The seating is stadium-style with twelve semicircular rows of fixed seats with attached desk tops. Each row is cut by two sets of stairs, one on either side of the room, with up to 26 seats per row. The lecture room has a maximum seating capacity of 255 students, and is usually 80%-85% full during any given lecture. Students enter the room through two doors at the top level of the stadium, facing wall-mounted boards on the lower level in the front of the classroom. The instructor has three chalkboards across the front of the room. Each chalkboard can be covered with a projector screen. The side boards have overhead projectors able to project acetate overhead slides. There is a computer and other multimedia material connected to an electronic projector for the center board only. The front of the room also has space for demonstrations between the first row of desks and the chalkboards.

The computer-based laboratory is a smaller classroom with tables arranged in a modified “c” pattern around three of the four walls and extending toward the center of the room (see Figure 1). This pattern allows the laboratory instructor to monitor the computer work of all students from a central position. This also removes distractions when the instructor wishes to communicate with the class, since all of the students must turn away from the computers in order to engage in the instruction. Groups in the laboratory are limited to pairs, but the configuration of the computers makes it possible for up to three pair groups to communicate with each other.

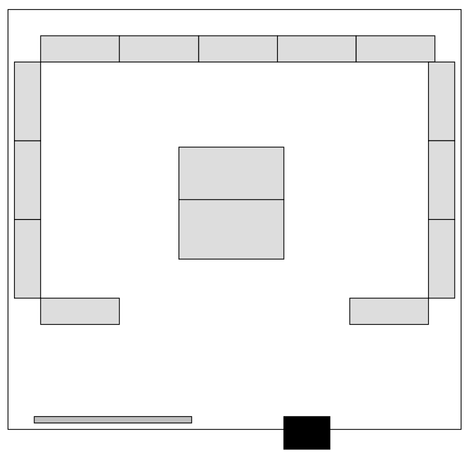


Figure 1 Modified "c" layout of the laboratory classroom

Instruction

My instruction model is based on scaffolded instruction. Specifically, I developed a new instructional system based on a mixture of the successful methods discussed previously. My final instructional model has as its major theme the inquiry cycle as developed by White and Frederiksen (1998). Within that major theme, different specific forms of instruction were developed depending on the content. Those instructional forms were Peer Instruction (Crouch and Mazur, 2001), concept maps and tutorials (Zeilik, 1998; Prather *et al.*, 2004), advanced organizers (Mayer, 2001), demonstrations, and laboratory work (White and Fredericksen, 1998). Before I talk about the major theme of inquiry, I will discuss each of the instruction techniques in more detail: Peer Instruction (Crouch and Mazur, 2001) as a concrete example of scaffolded instruction; concept maps and tutorials (Zeilik, 1998; Prather *et al.*, 2004) as they are explained in the literature; and advanced organizers as a thematic organization technique in the class. The laboratory work will be explained in detail in my discussion of our inquiry cycle based in White and Fredericksen (1998). I will not discuss demonstrations in detail, since these are standard physics and astronomy demonstrations that are discussed in great detail in the physics education literature (McDermott and Redish, 1999).

Peer Instruction.

As an example of scaffolded instruction, Peer Instruction (PI), developed by Eric Mazur, professor of physics, engineering, and applied sciences at Harvard University, includes lectures designed around short presentations highlighting specifics of the subject, each followed by a conceptual question where students take 2 to 3 minutes to formulate individual answers. The instructor evaluates these answers to determine whether the majority of the group has understood the concept. If there is significant misunderstanding (greater than 50%) the instructor may choose to intervene at this point and discuss the topic again. If between 50% and 70% of the students show understanding, small groups of students (PI) enter into a structured 2 to 4-minute discussion in an attempt to reach a consensus on the answer and convince others of the soundness of their reasoning. During this time the instructor moves about the room listening to the discussions and urging students to pursue their line of thinking. The instructor calls the class together, polls groups for their answers and underlying reasoning, then provides a summary explanation of the answer and moves on. Less is covered in lecture but it is covered more effectively. The final component of implementing a PI class is to have students complete a reading assignment before lecture that is closely aligned with the lecture content. Students are given both incentive and guidelines for thinking about the reading content. Students receive credit (5% of course grade) for completing a three-question, Web-based assignment over the reading. Two of the questions probe difficult aspects of reading and the third asks for what they found difficult or interesting. The instructor uses these responses to prepare for the lecture which overall saves preparation time and focuses content on concepts the students found problematic (Crouch & Mazur 2001). In this project my version of Peer Instruction included all of the aspects of PI explained above except the Web-based questionnaire.

Conceptual Astronomy and Astronomy Tutorials.

Conceptual astronomy instruction, developed be Zeilik *et al.* (1997) at the University of New Mexico, specifically targets the large lecture course with no laboratory section. His approach is based on conceptual understanding of astronomy content, with attention to cognitive studies in physics and psychology. Zeilik *et al.* (1997) developed instructional techniques centered around concept maps, small group tutorials, and a different type of assessment based on conceptual gains as seen in pre-post diagnostic tests. In developing instruction, Zeilik *et al.* (1992) began by identifying concept clusters. The concept cluster is a group of concepts that share a common theme in physics or astronomy. An example of a concept cluster is "Scientific Models". The results from Zeilik (1997) showed an increased mastery of content and concept pre- to post-instruction.

In my class, the concept maps were used during those parts of the course where there were many branches in the scientific model. For example, the mass of a star determines the life of that star. However, the early life stages for stars are all very similar. The later stages of life can be very different for stars of different masses. As an instructional tool, I give the students a pre-designed concept map with main ideas already encoded, but with plenty of space for students to annotate. The map that is built over the course of two weeks fills two normal sheets of paper when finished, but it shows the connections and directions of each step in the evolution of a star.

Conceptual astronomy and the tutorial approach to astronomy instruction (Adams, Prather, and Slater, 2005; Prather *et al.*, 2004) have a similar genesis. Both come from institutions that do not offer a lab class for the introductory astronomy course. Both use tutorials. Together they make complementary approaches to teaching non-major introductory astronomy.

Adams *et al.* (2005) and Prather *et al.* (2004) have developed a set of lecture tutorials for use in a large lecture non-laboratory introductory astronomy course at the University of Arizona. The lecture tutorials are designed to actively engage students in the topic of the day, with questions posed in a series of increasingly difficult steps akin to scaffolded instruction. In his study of the tutorials Prather (2004) showed that in a lecture-only astronomy class, significant conceptual gains can be made by introducing the lecture tutorials into the standard lecture classroom as described earlier. Although

Prather *et al.* (2004) mention the work of Redish *et al.* (1992) and diSessa (1988, 1993), the tutorials place more emphasis on the learning of content and concepts, but less time on how we construct knowledge from this content or these concepts, or how they become part of the general science discussion. That is to say that Prather *et al.* (2004) set out to develop conceptual understanding in the students, and they were successful. In my instruction, tutorials were used to help facilitate discussion. I viewed lecture tutorials as advanced organizers for my labs, and so I integrated them into the pre-labs.

Advanced Organizers.

It is Monday at 1:00 PM. The students are in their chairs for the lecture. I begin by asking some PI type questions that review the previous week and help to identify important aspects of the reading for that day. This takes about 10 minutes. I begin to talk about the lab for that week by a mini-lecture (15 minutes at most) that highlights the concepts and content for the lab that week. I give the students an organization to frame the work they will do in the lab and what we are expecting them to learn in the lab. I try to be explicit as possible in my discussion. I finish the day with a few more questions that incorporate some of what I have been talking about. Fifty minutes has passed and the students leave the classroom.

In this previous passage I stated that my Monday lecture included a discussion of content and concepts that the students will be seeing in the lab that week. This creates a framework that, as Mayer (2001) states in his book on educational psychology, helps the student “organize and interpret new incoming information.” This type of instructional format is known as an advanced organizer. Advanced organizers can take many forms: chapter summaries, introductions, key terms, outlines, or chapter questions. In the passage above, my advanced organizer took the form of a small lecture as an introduction that concentrated the students on the concepts covered in the lab.

Mayer (2001) explains that advanced organizers may supply "prerequisite knowledge" or even help in "making connections between incoming information and prior knowledge." According to Mayer (2001) there is evidence that advanced organizers create gains in retention of material when students lack prior knowledge and when the organizer itself is a concrete model (*i.e.* an illustration or example like my passage at the beginning of this section).

In my course advanced organizers played a large role. Take for example my concept maps for stellar evolution. They also acted as advanced organizers for the topic. Peer Instruction and advanced organizers were the instructional tools I used the most.

I have just sketched the instructional techniques that I incorporated into my final instructional model in SPRING 2007. In FALL 2006 the only advanced technique I used in the lecture hall was Peer Instruction. The WINTER 2007 lecture instruction consisted of all of the previously explained instructional techniques, but there was no thematic connection between the lecture and lab. By the final term, all of the previous instructional techniques were overarched by the theme of the inquiry cycle as developed by White and Frederickson (1998). In each case I used varying instructional techniques where they made sense (see my discussion of concept maps), but those techniques were always chosen to suit the various topics in the lab exercises. The next step in describing the design process is to discuss the inquiry cycle.

Inquiry.

The word *inquiry* when used in the discussion over education takes on three distinct meanings according to Flick and Lederman (2004). The first meaning of inquiry is the “fundamental principle on how science is conducted” (Flick and Lederman, 2004). Students and teachers might automatically associate this with the building and testing of hypotheses (Marzano, Pickering, and Pollock, 2001). The second meaning of inquiry is the various processes and thinking that take place in “support of the development of new knowledge in science.” In the framework of Hammer (1994) this might include someone who thinks gaining knowledge in science is a matter of applying the formulas to every possible situation, a naïve view of this form of inquiry. The third meaning of inquiry that Flick and Lederman (2004) distinguish is that inquiry “also refers to knowledge *about* the processes scientists use to develop knowledge, which is the *nature of science itself*.” My lab exercises built using the White and Frederiksen (1998) inquiry cycle try to incorporate all three meanings of inquiry as explained below

The inquiry cycle developed by White and Frederiksen (1998) is presented in Figure 2. They developed their ThinkerTools curriculum from the premise that students can learn how to learn in the classroom: metacognition can be learned. In this case, students can reflect on their own learning process to help them learn subject matter. Also, White and Frederiksen (1998) found that teaching students how to learn was beneficial to the low-achieving students. Most importantly, students do not have trouble learning science because of age or intelligence. Students have trouble learning science because they do not know how to construct conceptual scientific models, nor do they know how to reflect and monitor their progress while learning to build these models. To realize these ideas, they developed a process by which the students engage in the inquiry process as active participants in the classroom. Student thinking is scaffolded in two different ways. First, all students are asked to reflect on their findings and the findings of others, thus teaching the students metacognition. Second, student thinking is scaffolded by managing the inquiry process. This is done through the initial simplification of topics and a controlled introduction of more complex material through the cycle. The students learn to appreciate the process of inquiry by being involved in inquiry.

The inquiry cycle developed by White and Frederiksen (1998) can be characterized by the learning goals of each stage (see Figure 2).

- QUESTION: The students should formulate a question to address a topic. The question must not only be well-formed, but it must be something that can be investigated. If students cannot investigate the question then inquiry cannot proceed.
- PREDICT: Students should generate competing “hypotheses and predictions about what might happen and why it might happen.”
- EXPERIMENT: Students should design an experiment that can determine what actually happens. The students should be able to carry out these experiments.

- MODEL: Students should construct a model based on their data that includes pertinent scientific laws.
- APPLY: The students should apply their model to a new situation. This will help them investigate the applicability of their model, and give an indication of the limitations of their model. The limitations that students find can be discussed and investigated (see QUESTION).

During the course of my introductory astronomy class, inquiry takes on a new sense. The non-major introductory astronomy course is a survey course. In one sense that is a positive in that I am not bound by the amount of material to cover. But in another sense the survey nature of the course implies to the students that we will be covering a little bit of most of the material in astronomy. It would be encouraging to be able to spend the entirety of the course on one construct as White and Frederiksen (1998) do, but that is not feasible. Thus, I do not have students who can take the time to build and test their hypotheses. I must construct exercises that take them through the process. This means that my inquiry cycle is a modified version of White and Frederiksen (1998).

In the classroom and on pre-labs, questions guide and elicit responses that will build the hypothesis with the students, so in the prediction stage of the cycle students do not need to develop competing hypotheses, but merely to predict what answers they expect to get. Each student will build a similar hypothesis and we can test that hypothesis in the lab.

In addition, I have added a piece to the inquiry cycle which is OBSERVATION, and comes before question. This can be seen in Figure 3. Instead of the limitations of the model leading to new questions, the limitations may lead to new observations. However, the limitations may lead to a new question as well. My labs entered the cycle from both the observation section and the question section.

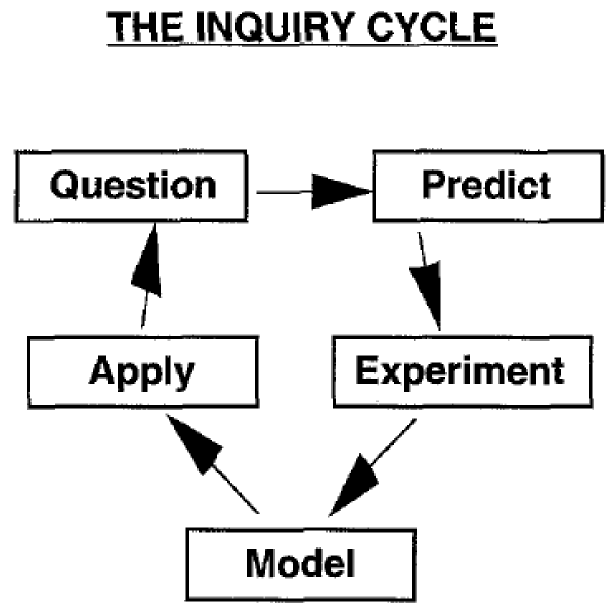


Figure 2 Inquiry cycle as it appears in White and Frederiksen (1998)

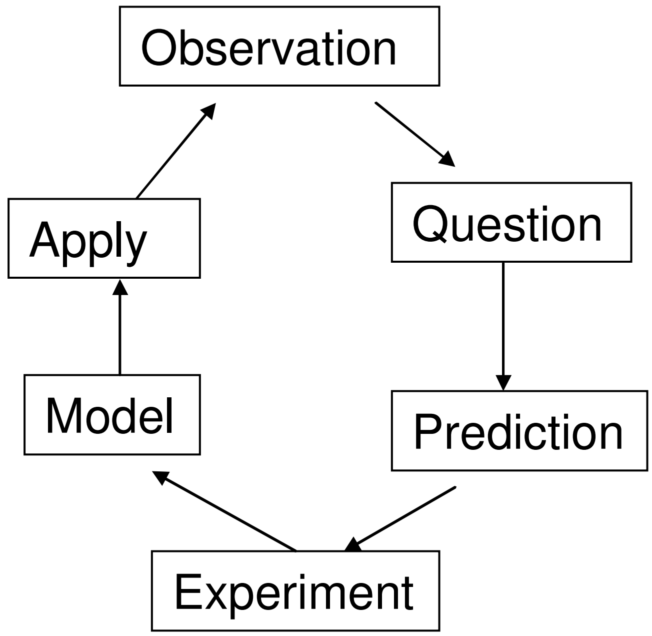


Figure 3 The modified inquiry cycle that I used to design instruction

The following instructional passage describes an application of my modified inquiry cycle:

It is a Monday afternoon at 1:00 PM and the lecture hall is full. I begin each class with an announcement. In this case there is a midterm coming up, so the students are told where to be and when. After the announcement, we begin a short reading quiz. All quizzes are taken using electronic clickers that interface with a laptop computer interfaced with an overhead projector. The laptop computer is also how I present my material. As mentioned by Mayer (2001), concrete examples are good forms of advanced organizers, so I attempt to show students as many contextual images from astronomy as possible. The quiz and announcement section of the day takes about ten minutes.

After the initial quiz, I make sure that each student has brought the pre-lab which will act as a tutorial and the first half of the inquiry cycle. The theme for this week is finding the mass of an astronomical object. Thus the pre-lab takes the student up through the Model stage of the inquiry cycle. However, I will only take the student up through the Experiment stage of the inquiry cycle. Using the inquiry cycle in this manner engages the students in thinking about the material that we discuss in class, and it acts as an organizer to emphasize the important aspects of the theme.

I begin with a slide that is the advanced organizer for this week (see Figure 4). This slide activates prior knowledge by attending to “stuff from before”. All of the "stuff from before" is thematically tied to this week’s material. This slide also has a list of terms or discussions that students should be “on the lookout for”. A version of this slide is in Figure 4. This slide will accomplish two goals of an advanced organizer, by activating prior knowledge and by helping a student organize “new incoming knowledge” (Mayer, 2001). A student will be thinking about how the "stuff from before" connects to the "stuff this week", and should understand the "stuff to be on the lookout for”.

Next we enter the inquiry cycle proper. The theme of this week is the masses of astronomical bodies. Since this topic is not as directly observationally-based as previous topics, I begin by posing a Question in the following manner: “In our text or on the Web or even previously in this class we have talked about Jupiter as being the most massive planet in the Solar System. We have a value for that mass, but how did astronomers figure that out? And can we figure it out from the stuff we have done up to this point?”

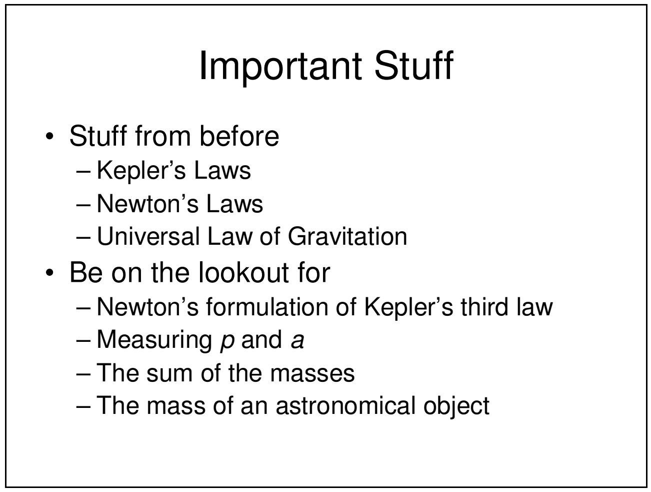


Figure 4 The advanced organizer slide for this passage

While in the inquiry cycle, I use the terminology of the cycle with the students for each step. In the case from above I write a slide that says Question. If there is no observational component that leads to a question, I let the students know that we will need to make observations, but they will become the data collection step of the experiment. This lets them know that an observation is not only the beginning step in much of astronomy, but an observation becomes part of the experimental process. As White and Frederiksen (1998) discuss with their inquiry cycle, the students have a guide that shows them direction, and it serves to direct the discussion during Monday and Friday.

After I pose the Question, I pose the same question to the students each week. “What ideas from the previous material or from the reading for this week could we use to answer this question?” Here is where I post an electronic question for the class. The question is “Which of these will help us find the mass of Jupiter” (see Figure 5). This allows the students generate alternative hypotheses without veering off of the theme of the week. This works to solicit the beginning of the Predict section of the inquiry cycle. In this instance we are not predicting about what might happen, but we are predicting which previously understood models will best help us to proceed with inquiry. It almost looks like we are finding a Model to apply, and that we are doing no predictions about what is going on. But in this case I have modified the instruction and broadened the definition of Prediction to be able to answer Questions like "How do we know?" This serves to show the students an authentic science experience, but also shows them how astronomers arrive at the "answers" we do without being able to go out into the universe and manipulate the object. This is the greatest modification to the White and Frederiksen (1998) inquiry cycle: the definition of Predict has been broadened.

Which of these will help us find the mass of Jupiter

1. One of Kepler's Laws
2. Newton's formulation of Kepler's third law
3. The universal law of gravitation
4. One of Newton's laws of motion

Figure 5 Hypothesis Slide

After students have reported their answers, I show a histogram of answers to the entire class. On many occasions this generates instant conversations among the students. When the students see the most popular answer and it is not theirs, they often think that they are "wrong". Many times in the classroom during these interactions, I will have a hand raised or a student speaking up. A typical interaction may develop like this:

- STUDENT: "I thought that it would be number 3, but most everybody answered 4. Why is that one right?"
- ME: "Did I say that one of these is the right answer?"
- STUDENT: "No, but everybody answered that one, so is it right?"
- ME: (To the class) "Show of hands, how many think that since this is the most popular answer, it must be right?" *Some hands go up, but many stay down.*
- ME: "We are predicting what we think will be the best way to proceed. At this point we need to decide which data we need and what types of observations will get us that data. There are multiple ways to proceed, but some are more economical than others. One method may not generate the answer I am looking for. But another method may require a billion dollar rocket, while a third method may only require me to take pictures of the thing in the sky with my telescope in my backyard. So in this case the only way that wouldn't work is the one that doesn't give me the answer I am looking for. The others may be valid, but we want the one that is most economical.

It is usual that due to time constraints we cannot spend an adequate amount of time on this point, but we move on. I tell each student to enter their prediction on the pre-lab. Since this is not a matter of right or wrong, there are no stakes. The only thing they are told is that they cannot change that prediction, but they will be allowed to write down a different prediction later.

The next step is to go through all the alternatives and discuss them. I focus my next five minute mini-lecture on discussing each of the alternate methods and what we would need to know in order to make use of that method. At this time the students should be thinking about their choices and weighing them against the alternatives. Our goal is to gently guide the students to the desired experimental procedure. After my mini-lecture, I post the same question again, as in Figure 5. Some students may stick with their first answers, but many choose using Newton's formulation of Kepler's third law, which is the method that we normally use to find the mass of an object. Newton's formulation of Kepler's third law tells me that I can find the sum of the masses of two bodies orbiting each other if I know the period and radius of the orbit. In this case I can use Jupiter's moons as the bodies in orbit around Jupiter. The question that might arise here is how then do I pull out the mass of Jupiter from the sum of the mass of Jupiter and one of its moons? This comes up in the classroom discussion. When students make their second prediction they are asked to respond to a section of the pre-lab that asks, "After reviewing the alternates, which method would you choose now?" They are then asked whether they changed their minds from the first prediction and if so why.

Once we have determined our method, we go into the Experiment stage. We need to ask what data we need, and how we can obtain it. In this case I will bring in data that I have generated along with images of Jupiter that shows how the orbits of its moons would look to us in the sky. The time constraints placed upon us require that in this section of the cycle the students take a role of observer while I model the processes for them. In this case, model means showing them the procedures that we would take in this experiment. When they apply our model (from the inquiry cycle) in the lab, they will be able to become the active participants.

The experimental procedure in this case consists of taking the data and running it through the calculations. This may take the form of me using the chalkboard, but normally I can develop a set of slides that will be as enlightening. Figure 6 shows a subset of the slides that would be used in this instance. Students are working along with me in their pre-lab. Once we have an answer, my work with them on the inquiry cycle is essentially at an end. This is where students get to pose questions or concerns about what we have just gone through. Each student will be responsible for applying our discussion by developing a Model (formula or qualitative statement) to apply in the lab.

During the time after I work through the Experiment with them, the question always arises about the sum of the masses versus the mass. After I am done calculating the mass of Jupiter, I show the students the accepted value from research. I then take them through a discussion on the difference between our answer and the accepted value. This shows them a concrete example of what scientists mean by error. I show that our error is small, so we can trust this method to get us the mass of Jupiter. "But then what happened to the mass of the moon?" will be the question that at least one student will ask. This is where we discuss relative sizes of Jupiter and its moons. By analogy, I talk to them about the Sun and Earth. The Earth is 1/300,000 the mass of the Sun. If I were to use this method for the mass of the Sun using the Earth, I would get the sum of the masses as the mass of the Sun plus 1/300,000 the mass of the Sun. That is like comparing 1 to 1.000003. They are different, but not very different. So this method would give me the mass of the Sun and I could ignore the mass of the Earth. I intentionally do not give them a direct answer for the Jupiter question, since this is the challenge that will help them build their Models to Apply.

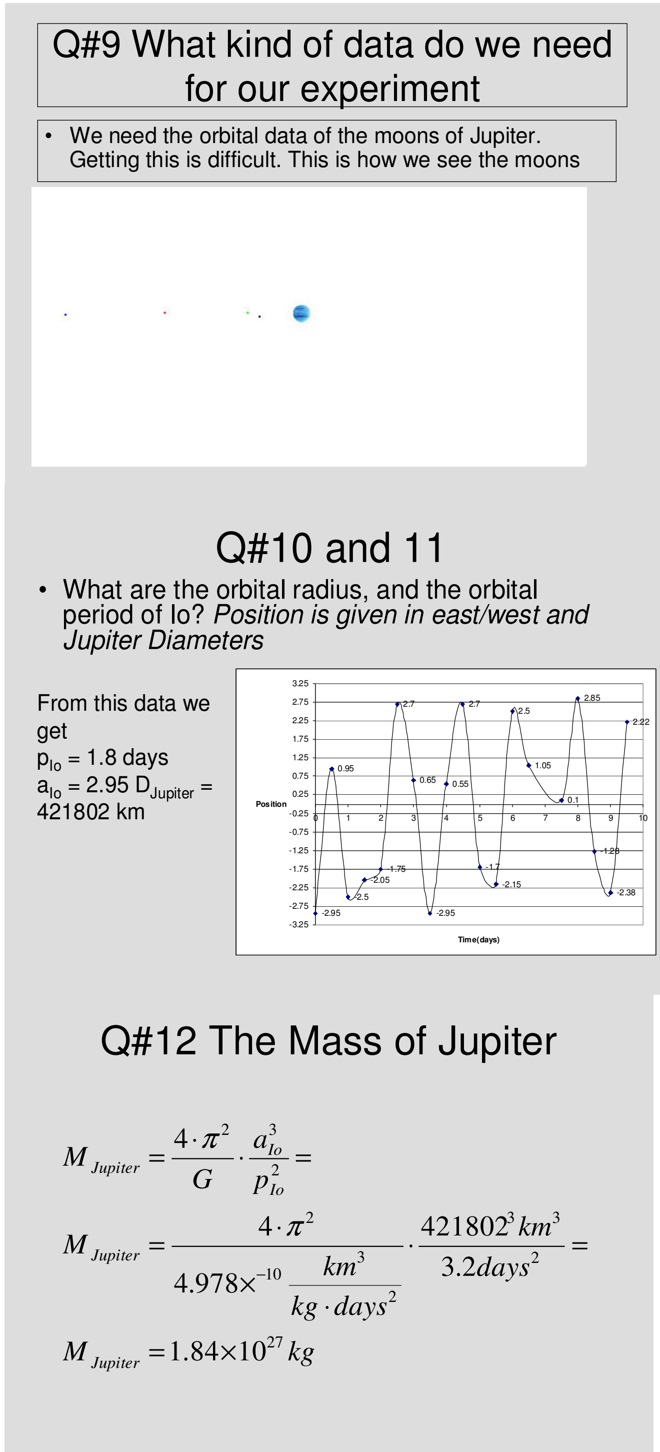


Figure 6 Experimental procedure slides

The rest of the inquiry cycle, applying the result, is completed in the lab. In our example of the mass of an astronomical object, they are applying the model we built on Monday to the mass of Uranus. This becomes a novel situation for students, since the images they will be working from look down on the orbit of the Uranus's moons. The perspective has changed and they need to think about the perspective.

In a typical lab section, the lab instructor will introduce the lab and connect it to Monday's material, by raising the idea that having built a model, now we want to apply it to see if it is usable in other situations. The lab instructor is a graduate teaching assistant (GTA) assigned to the class by the chair of the physics department. The graduate students are asked to pick which course assignments they want, so the GTA is in the classroom by choice. In order to train the GTAs to instruct the course, I followed my own method. Each week, either Friday or Monday, I would meet with the GTAs. The first portion of the meeting would be to discuss any problems that they saw from the previous lab, and changes that could be made to lab instruction. I would then have the GTAs get together as a group of three and work through the lab. The most junior GTA would lead the work, and the most senior would help when needed. As the GTAs worked through the lab, I would stop them at certain places and ask them to see how a novice introductory astronomy student could get confused by or misunderstand a passage. I would then ask them how they think they would go about helping the student understand what was going on. The GTA was never allowed to say, "I would give the student the answer." If the GTA had been with me before, he would talk about what he has done that was successful. If the GTA was new, she might respond with an intuition on how to help students learn. I would ask the GTAs that had been with me multiple terms whether they had tried it and how it had worked. This would proceed until the end of the meeting. My goal in the meeting was not to have the GTAs finish the lab entirely. They were expected to complete the lab once, but not necessarily with me. What I did expect was for the GTAs to communicate about teaching the material and what sort of questioning methods and interchange methods might work. In effect I was teaching the students about scaffolded instruction without explicitly calling it scaffolded instruction. This is an outline of the preparation of the instructor that the students would work with in the lab. Now we will continue describing the instruction of the students.

Once the GTA has introduced the lab and connected it to the previous material – another use of the advanced organizer – the students begin a round of data collection and analysis to apply the model to a new situation. The collection and analysis of the data is done in a prescribed manner. There may be computer images that students need to manipulate with a certain piece of software, or there may be already-generated data from which the students then collect values. In both instances the students are being challenged to take ownership of this data, through questions like "What do you think your data is implying?" or "Does your data look like what you expect?" Usually the analysis of the data consists of using spreadsheets, either to perform time-consuming hard-to-master calculations, or to display the data graphically for students to interpret. GTAs monitor the progress of the data collection and analysis and help where needed.

After the students have collected their data and analyzed it, they are asked to interpret and critique it. In the specific lab, the students are asked to find the mass of Uranus using each of four of its moons. This demonstrates the concept of accuracy for students. A question then is, "Why do you think we decided to find the mass using four moons and not just one?" This question is purposely vague enough that some students answer, "To make sure we are in the lab the full time". Mostly the question generates a communication between the lab partners and the GTA. The GTA will guide the students through to the concept that errors in the result can be minimized by taking and analyzing more data. A guiding question might be, "is there anything "weird" with your values for the mass?"

The final set of questions in the labs is designed to solidify the concept under discussion, but to also elicit thoughts about how science works. For this particular lab those questions are:

- The accepted value for the mass of Uranus is (*mass would be here).* How do you think your mass compares with the accepted value? *We are directly showing students the negotiations that take place in science. There is an accepted understanding and we compare with that. We never say the correct mass or the correct value.*
- What study could you come up with that would help you close in on the accepted value? *This elicits the idea that we test and then retest in science to become more accurate.*
- What other applications might we find for the model we applied today? *This is intended to elicit responses from the students for possible other avenues. The students should be beginning to see beyond the closed door of the class.*
- In those situations listed above, which ones would require that you need only modify your model a little? *One of the overarching concepts in the course is*

> *the usefulness (or lack thereof) of scientific models. Each model is expected to be viewed through the lens of a scientist who would ask where the model is sufficient as it is, where it will function with small change, and where it will not be useful at all.*

The final stage of instruction takes place on Friday. Students meet in the large lecture hall, and we begin with slides similar to the opening on Monday. Instead of questions, I launch into a mini-lecture of no more than 15 minutes, where I present the underlying theories behind everything we have done that week. Once my mini-lecture is complete I ask for any more questions from students. Then in PI form I pose a question about the topic. This initial question is always to identify the model that we manipulated this week. This is a way of summing up in the student's mind what all of this week has been about. The PI format of Friday's instruction continues with me asking a question intended to connect this week's content with the material covered in previous weeks, followed by small focused discussions about the answers. The last ten minutes of the class is devoted to previewing the topic for the following week. In this specific case the following week begins a new set of material not directly connected to this week's material. This week was in the Solar System, the following week will be about stars. So the preview of the following week is a question without answer, "How do we classify the stars". Thus we have gone through a complete one week inquiry cycle. The next week begins at the top of the cycle. In some instances we are expanding on the current week's material. In other cases, we are changing topics completely.

The previous passage shows the inquiry cycle with references to where I deviate from White and Frederiksen (1998) and what my expectations for student thinking at important stages. In this passage, I connected loosely to my epistemologically-based instruction framework. I will make this connection more concrete here. My framework for student epistemologies is based on a framework of Hammer and Elby (2003), Louca *et al.* (2004), and Schommer (1990). In my framework, students have some coherent belief system, but that belief system is situated in the context. This means that in one context, a student could have a sophisticated belief in one context, but that very same belief could be naïve in another context. To inform my instruction I employ a broader view from Hammer (1994) where he identified three epistemological constructs in the students he interviewed (Table 3). In this case we can construct instruction built to address these three belief structures specifically through our scaffolded instruction in the classroom.

Table 3 ***Range of beliefs from naïve to sophisticated epistemologies about three aspects of physics knowledge.***

Belief	Naïve	Sophisticated
Structure of physics knowledge	A collection of isolated facts and pieces	A single coherent system
Content of physics knowledge	Formulas	Concepts that underlie the formulas
Learning Physics	Receiving information from the authority	An active process of constructing one's knowledge

Let us apply the three beliefs in Table 1 to analyze the instructional unit on planetary masses described earlier. My initial question about the mass of astronomical objects is: how do we know and how do we find out? This seems a straightforward question, but in the epistemological framework from Hammer (1994) I am directly addressing the belief that knowledge comes from authority. "There are these values in the book, but I don't need to take them on face value, I can actually find them out myself and I can confirm the values if I choose". Furthermore, by explicitly showing the student that the value is not a "mysterious" calculation, but stems from an underlying physical principle, I have contradicted the belief that scientific knowledge is a collection of facts. Finally, the instructional example also demonstrates that solving physics problems is a matter of understanding those underlying concepts and not just manipulating formulas. Thus all three major belief structures have been addressed.

Not only can I argue that this instructional model addresses the major beliefs found by Hammer (1994), but this instructional model stimulates development in the students' epistemology of science through many of the summation questions at the end of the lab. Questions that ask “how does your value compare with the accepted value” are intended to show the tentative nature of science. This question elicits the idea that there is an accepted value which is open to change in the course of investigation. Our labs also show the empirical nature of science, in that we base our knowledge on observations of the natural world.

The previous description of a typical week is an example of the final implementation of the instruction in the non-major introductory course. According to table 1, this complete model was not implemented until SPRING 2007. In the rest of the paper I will refer to that example with annotation to discuss the findings from the epistemological measurements that I made. I chose to implement instructional augmentations in a staged progression in order to study the effects on the students of certain components of this instructional method. During the break prior to each implementation, I reviewed the post-instruction epistemological measurements to see if there was a change from the prior term. This did not change the implementation schedule. What it affected was the presentation of the material. If the post scores showed a negative trend, I reevaluated the instruction materials to identify what I could change to foster positive growth. This project initially stemmed from another smaller project intended to study the epistemologies of the non-major introductory science class. This small study for internal use provides baseline measurements, so I will introduce the baseline results prior to introducing the implementation of the course and the results therein.

The Initial Epistemological State of Students

I had been interested in understanding how students in the astronomy course, where I had been a GTA, viewed science and building knowledge in science. I developed the initial framework that would guide my research for this course through the works of Hammer (1994), Hammer and Elby (2003), and Elby (2001). Mostly I concentrated on the work of Elby, because it seemed that his instrument, the Epistemological Beliefs Assessment for the Physical Sciences (EBAPS; Elby, 2001), would work best. My final choice of instruments is explained in the discussion on the EBAPS in the following chapter. EBAPS is a 30 question forced-choice Likert-style instrument. It is designed to measure student epistemologies of science, mostly focusing on the learning of science when one reads the questions. The following discussion explains the procedure that I used for this baseline study.

The structure of the course that students attended was similar to that described in the Setting section above. Students attended three fifty-minute lectures and one two-hour laboratory session each week. The laboratory exercises were intended for students to apply their knowledge gained in class or the readings. Many lab exercises were scheduled at times unrelated to the material covered in lecture, so that activities in the two contexts did not complement each other. Student evaluations reflected this disconnect with many comments implying “I wish the labs had more to do with the lectures." The lectures were didactic presentations with students watching and taking notes on a 50 minute PowerPoint presentation given by the instructor (not the author). There was no reform-style instruction taking place in the classroom or laboratory.

In spring term 2004 the students were given the EBAPS as a pre-and post-instruction survey. The students took the EBAPS with them from class, and were required to turn them in to a secured location within one week. Students filled out a scantron sheet and returned it as well. Removing students that did not wish to participate or that did not turn in both pre and post surveys left 108 participants. The pre-instruction EBAPS scores for the students had a median value of 2.51 (SD = 0.37) overall on a 4 point scale, so the students are slightly more sophisticated than naïve on the EBAPS measures.

The overall median for the EBAPS post-instruction in Spring 2004 was 2.02 (SD = 0.37), a significant decrease from the pre-instruction overall score, $t(214) = 9.738$, $p = 0.0001$. In my discussion of the MPEX study, in the opening chapter of this document, we saw that Redish *et al.* (1998) found the largest decrease pre-post in the MPEX data for those physics classes that did not use any reform-style instruction. My initial findings in spring of 2004 show that this problem plagues not only physics classrooms, but also non-major science classrooms that employ traditional teaching methods.

This initial observation set me on the path of the research project discussed here. The design and implementation of the course is directly motivated by the decrease in student epistemologies as measured by the EBAPS. However, there are some limitations in comparing the spring 2004 to FALL 2006, WINTER 2007, and

SPRING 2007. I did not take any demographic data for spring 2004. I must assume that population was equivalent to the groups in my research project, a hope bolstered by later observations that the FALL 2006, WINTER 2007, and SPRING 2007 groups are equivalent. There are minor variations in the 2006-2007 groups, but overall they are very similar. The lack of demographic data for 2004 is an unfortunate handicap of this study.

The students taking the non-major introductory astronomy course at Oregon State University in a "traditional" lecture/lab course model suffered dramatic deterioration in their epistemologies as measured on the EBAPS. In the next section I will discuss the implementation and results from the EBAPS for each individual term of the revised instruction. I will then make general remarks about the reform project.

Instructional Design

The design of the labs began after Spring 2004. The first decision was what topics would be included. When introducing an instructional innovation, like explicitly teaching the nature and epistemology of science, a question arises: What is being left out? In his description of PI, Eric Mazur (1997) lets us know that the amount of material that he covered was reduced. The rule of thumb is fully one third of the amount of material covered in a traditional lecture class can be covered in a PI style class. This is due to increasing the depth of coverage of the remaining topics. Similarly in the astronomy course, we needed to decide what material we would no longer include.

The content decisions were made in three major steps. First, we defined and discussed the overarching themes of the course. This meant that we needed to specify the learning outcomes that we wished to see in the students. These outcomes were first generalized in the following categories:

- The process of science through inquiry; this was our major theme that would override all others. If at any time we were unsure about the inclusion of a content or concept, we looked at it through this lens. We were always asking, does this content support a sophisticated epistemology of science or foster sophisticated views on the nature of science? If the answer was no, then we generally discarded it. In this manner, we initially discarded at least 30% of the content previously covered in the course.
- The role of models in science; this was the second overarching principle. Whenever we could bring scientific models into the discussion, we would. We sought an engagement with a model, or the development (with help from the instructor) of a model based on evidence that we had looked at during the class or lab.
- Classifications in science; this is the third overarching theme. This theme has little influence on the content, since we discuss classification in most of the topics. Thus, this is not as difficult to design around as the previous two.

After the overarching themes were defined, we discussed the specific content of the course. The course content fell into three large sections:

- The Solar System: We determined that students should have a deeper understanding of the planets than was provided by the text in use in the previous course. We felt that to help students understand the role of classification in physics and astronomy, we would engage them in a discussion over the Solar System as a whole. We would look at similarities and differences and build a model of the Solar System and its formation.

- The Stars: The stars and their lives are one topic where students always show an interest. This is another place that students can develop an understanding of models and classifications.
- Galaxies and Cosmology: This topic is the least extensive but most difficult of the three. The students do not do as much inquiry into this topic as they do in the other two. This is for various reasons, but mostly it has to do with the unfamiliarity and abstraction of this material. However, this is the last topic covered in the course each term, so by the time we have reached this point the students should have an understanding of how astronomers have developed their understanding of the cosmos, and we can discuss the topics in this framework.

The second step toward content decisions was to inventory and evaluate prior material. This fell into two categories: what did we have in house to build upon, and what had others done that we could exploit in our lab framework? At Oregon State University, we had a number of lab exercises that had been developed in prior years. Collectively, they did not form a coherent whole, but subsets of the labs connected well with each other. We began with these connected labs as a skeleton frame and redesigned them to fit into the inquiry framework. These were then introduced to the students during Fall 2005. Each lab was followed and discussed on a regular basis with the GTAs and among the group working on the inquiry cycle. Strengths were built upon, and weaknesses were discarded until we had five labs that we felt worked toward our desired learning objectives.

Using the five original labs, we proceeded to look into exercises that others had designed. We found the Contemporary Laboratory Experiences in Astronomy (CLEA, Marschall, 2000). CLEA is a set of experiments done using a virtual observatory experience. Students log in to a virtual telescope and take synthesized data that they can then analyze. The experience attempts to be as close to authentic as possible. We vetted the CLEA labs through the same process as our own labs and concluded that only one of them would be useful for us in our course. We also looked at adapting the tutorials developed by Adams, Prather, and Slater (2004). We determined that these tutorials would be useful in the lecture portion of the course, but that the work to develop them into lab exercises was beyond our time frame.

In the end we developed three lab exercises from scratch. Of these three exercises, two were incorporated into the lab section of our course. This gave us eight lab exercises for the term. These eight new exercises were introduced in Winter 2006. They were each studied for content, time, and exam outcomes. Since this was a development stage and not the project, no data will be shown. Between Winter 2006 and Spring 2006, the labs were retooled. They were then studied in Spring 2006, and retooled one more time before we decided that they fit our learning objectives.

While designing and developing lab exercises, we were also designing lecture instruction that would become connected to the lab exercises. Lecture instruction was designed using prior instructional models. Those models are researched and tested in large lecture settings. We used Peer Instruction (PI) developed by Mazur (1997), conceptual astronomy developed by Zeilik (1998), and the tutorials developed by Adams, Prather, and Slater (2004). These are all researched instruction techniques that have been shown to increase student content knowledge.

We determined that the main focus of lectures would be content development that supported our overarching goals and the labs. The connection between lecture and lab became the main focus of design throughout the remainder of the project. The design discussions revolved around the following questions:

- What content do the students need in order to successfully complete the lab exercise? Definitions? Derivations? Examples?
- What prior content did we cover that we can reprise in the upcoming content? This is a case of keeping the connections going through the course. In order to show the students the connected nature of the scientific process, we should explicitly connect one topic to another.
- Which of our goals will this exercise help to meet? At the beginning of each lecture, we should tell the students the answer to this question. We shouldn't just ask it for ourselves.
- What instructional technique would best help us to accomplish the goal and give fitting answers to the previous question? Do we do an example like the Mass of a Planet above? Do we lecture blatantly? Do we work out a tutorial? In the end we decided that for the beginning of the week, the pre-lab discussion worked best. However, the rest of the week was split into

many different instruction styles. PI was used in every class, but the context it was used in changed.

The third aspect of the design was to determine an implementation schedule. We wished to know how just an inquiry-based lab would affect student changes in content and epistemology. In order to measure this we needed to stagger the implementation. We chose to implement the full lab schedule first. This means that even though I was working on lecture instruction material to include, I could not implement it for at least one term.

Implementing the labs without all of the other instructional material served another purpose. Data taken and discussions with the instruction team (instructor and GTAs) would allow us make relevant changes to the labs when we introduced them into the connected lecture/lab instruction model. In this manner we could identify how inquiry affects undergraduate non-major students. These findings will be discussed in another manuscript.

The following discussion will address each stage of implementation and what we discovered during and after each stage. A more detailed empirical analysis will be completing in a separate manuscript.

Implementation

This section will detail the implementation of the instruction methods during the project year of 2006-2007. Each term will be detailed and discussed.

FALL 2006

According to the introduction of the design section of this manuscript, the implementation took place in three phases. In the first phase, we introduced a new set of labs based on the inquiry cycle explained earlier. These labs included the entire inquiry cycle. Above we discussed a pre-lab component that took place in the class. This was in the final design implemented in SPRING 2007. All of the components of the pre-lab – observation, question, prediction, and experiment – were integrated into the lab in an introductory section. The introductory section of the lab was very short. At times only one question was devoted to each section of the inquiry cycle. Scaffolding was accomplished through the interactions of the students with the GTAs in the lab and Peer Instruction in the lecture.

The lecture portion of the course was done at its own pace. Students could be approaching content in the lab up to two weeks after the material was covered in lecture. The only instruction method use in the lecture was Peer Instruction (PI). The questions posted during lecture would use a set of purely conceptual questions and set of more epistemologically-based questions. An example of a conceptual question is "The reason that the Earth experiences seasons is…" An example of an epistemological question would be:

> Suppose that there are two competing models for the extreme tilt of Uranus's rotational axis. How would one go about confirming which one is the better model?"

A. We would ask a lot of professional astronomers to see which model most of them think is correct.

B. We would create computer simulations for both models and see which one leads to what we observe about Uranus today.

C. We would see if we can't find evidence that would indicate one model could not happen.

D. If we did all of the above, we would be building enough evidence toward the chosen model.

This question asks the student to think about how we gather evidence and about what qualifies as evidence. Questions like the previous are worked into the discussion throughout the term. When students are asked these types of questions they are given credit for answering only, and then we discuss in class why one answer would be more valid than the others.

The biggest instructional issue that we ran into with the course was during the lab sections. During our weekly meeting, the GTAs and I discussed the fact that many of the students were not completing the labs. This became a problem for every lab of that term. In the end I was required to weight the labs differently than I had at the beginning of the term. I carried the new weighting through the rest of the academic year. During FALL 2006 I was involved with rewriting the lecture section of the course, so I did not change the labs until SPRING 2007.

Student epistemological changes were measured with the Epistemological Beliefs Assessment for Physical Sciences (EBAPS) using a pre- post-instruction format. Their content changes were measured using the Astronomy Diagnostic Test (ADT) and exam averages. Both of these measures together constituted the only data collection done in this study. I will highlight the findings of the EBAPS as they informed instruction.

The pre-instruction median for EBAPS in FALL 2006 was 2.35 (SD = 0.55). This is lower than the mean for Spring 2004, but it is not significant, $t(172) = 2.83$ $p = 0.1$. The pre-instruction results for FALL 2006 still indicate that students enter slightly more sophisticated than naïve. Why that is so was not part of the discussion that entered into the design of this project.

The post-instruction EPABS median for FALL 2006 was 2.42 (SD=0.44). This is not a significant change from the pre-instruction median for FALL 2006, $t(278) = 1.42$, $p = 0.78$. However, this is a significant increase in the overall post-instruction median compared to Spring 2004, $t(278) = 1.92$, $p = 0.027$. It is apparent that just introducing the inquiry-based labs made a significant change in the post-instruction EBAPS scores from Spring 2004 to FALL 2006. The inquiry cycle did not degrade the students' epistemologies of learning science, nor did it increase their epistemologies of learning science. There were two issues that arose in my findings that I will discuss next.

The two major issues that we had in FALL 2006 were labs not being completed on time and lack of significant increase on the EBAPS. First, the lack of significant increase in student epistemologies indicated that the lab might not be enough to foster increases in student epistemologies. The next stage of the project included new lecture instruction which meant that I needed to concentrate on that portion. The next section discusses implementation in WINTER 2007; the issue with the labs was not addressed for WINTER 2007.

Another issue that I initially did not deem as major was the lack of questions devoted to the inquiry cycle in the labs. Post-instruction EBAPS scores indicated that the labs were changing the way that students viewed learning science. The design plan included the introduction of pre-labs in SPRING 2007, which would allow more in-depth discussions in the lecture hall. Those discussions would include the steps in the inquiry cycle as described previously.

In order to gauge the way that lecture instruction changed student epistemologies, I chose to leave all of the lab issues just discussed until SPRING 2007 when it would be important that these issues were addressed. The next section discusses the introduction of the new lecture instruction in WINTER 2007.

WINTER 2007

Phase two of the redesign was to incorporate varied instructional styles into the lecture section of the course during WINTER 2007. As discussed earlier, Peer Instruction (PI), defined by Mazur (1997), was our primary form of instruction for FALL 2006. PI was not used in every meeting in FALL 2006, since PI is not always the best fit for some content. In WINTER 2007, PI became a tool in the larger instructional model I was developing. Other instructional techniques described previously – conceptual astronomy (Zeilik *et al.*, 1997), tutorials in astronomy (Prather *et al.*, 2004), and advanced organizers (Mayer, 2001) – were also used as tools. The overarching theme became scientific inquiry in astronomy. Using a tutorial, advanced organizer, or concept map with the small groups discussing the various aspects of the classification could help complement PI.

In WINTER 2007 each week's goals were looked over to determine what type of instruction would best serve the content. This is what Wiggins and McTighe (1998) would call backward design. Instead of deciding on an instructional style and trying to fit the content objectives to that style, we decide on the content objectives and fit the instructional style to those objectives. Backwards design makes it easier to develop scaffolded exercises, since we can identify the skills needed to accomplish the learning objective and instruct students in those skills.

Labs WINTER 2007 followed the same format as FALL 2006, but some labs needed to be rewritten for typos and language. I chose to concentrate on language issues due to the jargon that enters into any science discussion. In many cases I omitted jargon for a longer simpler definition. In other areas the jargon was a better fit. An example of replacing jargon in the astronomy course is the types of supernovae. Astronomers use descriptions such as Type-Ia and Type-II. These refer to very different situations as to why the supernova happened, but they can be confusing to the students. A better description of a Type-Ia supernova is white dwarf supernova. This is longer but simpler and it describes exactly what has exploded. The term supernova itself is jargon, but it simplifies the discussion, where an alternate description would be more confusing.

This next phase of design was implemented in WINTER 2007. Similar to FALL 2006, the lectures in WINTER 2007 proceeded at their own pace with lectures and labs overlapping little or not at all. This disconnect is the main failing of the lecture/lab format for large courses. In this method the students really are in two different contexts. My impression is that because little is drawn from the labs in course evaluation, students ignore work being done in the lab. During my assessments I asked specific questions that started with "In the lab,…". These questions were designed to show the students the connection of the lab with course evaluation. This is not a confounding factor in this research, since that format was consistent in all terms of this study.

Administration of the EBAPS and ADT in WINTER 2007 was done on the same time scale as FALL 2006. Student took the pre-instruction EBAPS within the first week of courses. Time constraints made it difficult to administer the EBAPS on the first day, since I had to introduce the study to students in order to obtain their consent. A short discussion of the WINTER 2007 EBAPS findings is presented below.

The pre-instruction EBAPS median for WINTER 2007 was 2.55 (SD = 0.39). This is not significantly higher than the pre-instruction EBAPS median for Spring 2004, $t(279) = 0.63$, $p = 0.26$. However, it is significantly higher than FALL 2006, $t(343) = 3.88$, $p = 0.00$. The low median on the EBAPS for FALL 2006 will be inspected more closely in the next manuscript. What we see is that students consistently enter the course slightly more sophisticated than naïve.

The post-instruction EBAPS median for WINTER 2007 was 2.46 (SD = 0.49), not significantly lower than the pre-instruction median for that term, $t(344) = 1.6$, $p = 0.055$. There is also no indication that this is significantly different from the post-instruction median for FALL 2006 $t(343) = 0.64$, $p = 0.17$. It was apparent to me that varied lecture instruction had no further effect on students' epistemologies than the inquiry-based labs alone. What does this mean for our instructional framework?

The inquiry-based labs had the greater effect on student epistemologies when compared to Spring 2004. This is seen in the similarities in post-instruction EBAPS medians for FALL 2006 and WINTER 2007. In the context of the instruction, more scaffolded instruction could take place in the lab compared to the lecture hall. Students were in smaller sections, allowing more time per student. Questions on the labs required more reflection on the material just covered. The labs made coherent connections, while the lectures were self-contained going from topic to topic daily. This did not allow students to reflect on the material in any metacognitive way.

During WINTER 2007 we had issues arise that were similar to FALL 2006. Namely the students could not complete the labs in the two-our time frame. This became the main focus for the next stage of design. Students would need more time to reflect during the lab class. This would require that we do some of the inquiry cycle during Monday lectures. My main focus became the production of pre-labs that incorporated the first stages of the inquiry cycle. The implementation of the final phase of instruction will be discussed below.

SPRING 2007

The third and final phase of instructional redesign, SPRING 2007, was to introduce a pre-lab component that would be specifically discussed on Monday. Lecture and lab became a coherent whole where I introduced the inquiry on Monday and summarized each week on Friday. The instructional example described previously typifies the instruction used in SPRING 2007.

The major issue that the pre-lab solved was the time issue. Students were completing the labs in the two-hour time limit with fewer problems. Their questions could be probed in a deeper manner with the GTA, since the reflection questions were being reached with more time left in class. This should allow the students to become more metacognitive. This should reflect in their EBAPS scores.

The pre-instruction EBAPS median for SPRING 2007 was 2.43 (SD = 0.39) while the post-instruction EBAPS median for SPRING 2007 was 2.42 (SD = 0.39). In this case there is no change in the pre- to post-instruction EBAPS medians. The post-instruction median for SPRING 2007 was not significantly different from the post-instruction median for FALL 2006, $t(314) = 0.068$, $p = 0.47$, and WINTER 2007, $t(315) = 1$, $p = 0.16$. This leads to the next discussion.

In connecting the lecture to the lab, it was my intention for students to see the application of the material covered in lecture. Experience tells us that the traditional lecture and lab courses cover material essentially in a vacuum. The lectures proceed at one pace, while the labs proceed at another. In some instances material covered in the lecture will never be seen in the lab. In other cases the labs are either weeks behind or weeks ahead of the lectures. Students are left to make the connections to material that they may have not seen or seen several weeks ago. Assessment in the course will not explicitly use the lab work. Students will then come to see the labs as mindless exercises and not as learning environments.

The disconnect between labs and lectures persisted in FALL 2006 and WINTER 2007. However, in SPRING 2007, I explicitly discussed the labs in every lecture. Often I would bring in PI questions directly from material in the lab. These questions would be based on the GTA meetings when we would discuss the parts of the lab that students struggled with. It is apparent that this connection did not lead to a significant post-instruction change in students' epistemologies. Finer detailed analysis of all of the data collected was necessary to see the changes made to students. This is discussed in a subsequent manuscript. The next section of this manuscript is a discussion of the design and implementation of the redesigned instruction explained in this manuscript.

Results

Table 4 shows the pre- and post-instruction EBAPS scores for each term in this manuscript. We see the significant drop in post-instruction scores for Spring 2004, $t(214) = 9.738$, $p = 0.0001$. Also the post-instruction score for Spring 2004 is significantly lower than the post-instruction score for any of the terms in 2006-2007. This indicates that the implementation of an epistemologically-based instruction model helped. The lack of significant increases pre- to post-instruction over the three terms in 2006-2007 indicates that there are design issues that need to be addressed.

First, we see that the introduction of the inquiry cycle led to significant change in post-instruction scores in FALL 2006 compared to Spring 2004. As White and Frederiksen (1998) indicated, we should expect students to better understand the content if they are required to reflect on it. It appears that the labs required the students to reflect on the epistemological material at a deeper level. However, it doesn't appear that the lecture instruction had much effect beyond this.

Table 4 EBAPS scores for the four terms discussed in this manuscript. All post-instruction scores for 2006-2007 are significantly higher than the post- score for Spring 2004 ($p < 0.05$).

Term	Pre-instruction EBAPS	Post-instruction EBAPS
Spring 2004	2.50	2.02*
FALL 2006	2.34	2.42
WINTER 2007	2.54	2.46
SPRING 2007	2.43	2.42

* indicates significance at the 95% confidence level, $p < 0.05$

The correlation of the coherent instruction model where lecture and lab are considered as parts of a whole leads to some ideas. First, students see a connection of the general ideas discussed in lecture to specific contexts discussed in the lab. Second, they can see science as an evolving process instead of something that was figured out long ago. Third, students can see that there is a connection between scientific knowledge and their real life. This should lead to students that have more sophisticated epistemologies of science. However, we did not see this in the data which brings up two possibilities. First, we may not have time in a ten week course for students to make significant changes in their views. Second, we may have had an effect on a subgroup of students that we could not see looking at the entire student population. This second idea will be detailed in the second manuscript.

In the epistemological framework, it appears that students have a coherent set of beliefs about learning science, but that those beliefs are context dependent. In this case it appears that the lectures and the labs were two different contexts for the students. The lab context was one that fostered student epistemological gains. The lecture did not add to what the lab was doing for the students. This is seen in the evenness of the post-instruction EBAPS scores for the three 2006-2007 terms. Students must have approached the lectures in the same manner as they had prior to my instruction and ten weeks was not enough time for them to become comfortable in the different instructional style.

The comment above about comfort in the instruction style leads to a discussion about the post-instruction EBAPS scores. All students entered the course at similar epistemological levels. These are slightly more sophisticated than naïve. During the ten week course they were introduced to a new lecture method, one that they probably have not seen elsewhere, and a new method of thinking about science. At the end of ten weeks of lecture, students may be trying to internalize the epistemological changes that we wish to see, but they need more time. It may be that both the new instruction and new thinking that we are introducing to these students interfere to an extant that it would take students much more time to become more epistemologically sophisticated. Finally, the students may be spending more of their time studying content, since they may feel that content is what they are really trying to learn.

The previous discussion collected all of the findings and showed that there is an effect on students' epistemologies using my redesigned instruction model. The difficulty in moving the students beyond the post-instruction EBAPS results for the three terms in 2006-2007 may lie in the lack of time students have to internalize the instruction. Also, the students may focus on content, forgoing thinking about science for trying to figure out the seasons. In the final section of this manuscript I will make some general comments about this project and list some possibilities for future instructional changes.

Conclusion

Our goal as instructors of science is to develop, in our students, necessary thinking and reasoning skills. Through our instruction we want our students to gain content knowledge, but to also have a sophisticated idea of what it means to be a scientist and what it means to do science. Curricula and instruction in physics and astronomy have addressed the content part of our needs (Adams, Prather, and Slater, 2005; McDermott and Mazur, 1999; Prather *et al.*, 2004; Zeilik *et al.*, 1997). Recent efforts to measure students' epistemologies of learning science have been made in the physics course (Elby, 2000; Hammer, 1994; Hammer and Elby, 2003; Redish *et al.*, 1998). The same effort has not yet been made in astronomy where Zeilik *et al.* (1997) have argued it is needed.

I have shown that when an instructional model is developed to specifically address students' epistemologies, we can foster those changes. The important aspect of this instruction is scaffolding. Regardless of the instructional method, scaffolding was the overarching instructional technique that I used. The second important aspect of my instruction was the inquiry cycle modeled from White and Frederiksen (1998). These coupled together showed the best post-instruction EBAPS scores during this study.

Besides PI, concept maps, tutorials, and advanced organizers in my instruction model there are some other instructional avenues that could be explored in order to foster student epistemologies of learning science. First, the labs need to be more reflective and open ended than they are now. Asking students to design their own experiments within the boundaries of what can be done in the lab would lead them to a better understanding of the processes that professional astronomers go through to gain knowledge. Adding more reflection to the labs would encourage the students to be metacognitive, leading to more sophisticated epistemologies.

Finally, I am going to discuss what other research projects could stem from this project. Most of these would incorporate more interactions with the students during the term.

The first avenue of research that needs to be explored is the development of an instrument to measure student epistemologies. We saw from the discussion of the EBAPS that the overall Chronbach alpha was 0.79. The EBAPS is measuring a construct overall. However, the subscales that Elby (2001) describes have alphas low enough that we can infer that they are not measuring the same construct. In fact, one of the more important subscales – Real-Life Applicability – has an alpha of 0.04. This indicates that none of the questions are probing the same construct as the overall score.

The EBAPS measures students' epistemologies of learning science rather than their epistemologies of science. From the introductory chapter to this document, we see that it may be difficult to probe students' epistemologies of science, but we may be able to carefully construct an instrument that does this.

Another area of research that could be followed stems from the lack of post-instruction changes between the 2006-2007 terms. This is an indication that the lecture portion of the instruction was not as effective at changing student epistemologies as the inquiry-based labs. This research should tie to the creation of a second EBAPS.

This would involve audio, video, and interviews of students within the course. We would need to identify those students with naïve and sophisticated epistemologies. We could follow each and see when and if changes are happening. Also, this would give us an insight into which parts of the instruction are making the most positive impact and which parts are having a negative impact.

One final anecdote that comes from this new instruction is the number of students inquiring about additional science courses that they could take at the university. Before the introduction of the new curriculum it was the rare student that wanted to know more about other science courses he/she could take. After the curriculum, I have at least ten students up to twenty students that wish to take other science courses in either physics or some other science discipline. This is anecdotal, but it implies that I am fostering students' epistemologies about learning science.

CHAPTER 3

Effects of Inquiry-based Astronomy Instruction on the Epistemological Beliefs and Content Knowledge of Non-Major Science Students

Matthew F. Price

Abstract

Research on the changes in student scientific epistemologies in the introductory physics and astronomy classroom has indicated that even in the more successful reform physics instruction and curricula, students do not change their views about the structure of science (Redish, Saul, and Steinberg, 1998). However, Elby (2001) showed that instruction in a physics course intended to foster epistemological changes will lead to students thinking about physics in a manner similar to professionals while maintaining content acquisition. Our study uses results from a non-major introductory astronomy course, redesigned to foster student epistemological changes, to answer the following questions: a) can an epistemologically focused non-major course foster changes in student epistemologies; and b) what changes in student content knowledge occur after epistemologically based instruction in the non-major introductory astronomy course? Introducing a progression of epistemologically motivated instructional enhancements in winter and spring, we found that the incidence of negative epistemological outcomes decreased in successive terms. This decrease in epistemological losses from fall to winter to spring cannot be attributed to any measured demographics, implying that it is correlated with instruction. A group of students with especially large epistemological gains in the fall, not evident in winter or spring, is identified as a group of especially epistemologically naïve students entering the fall course. We also found that students with mature epistemological beliefs had similar content outcomes regardless of instruction methods. Also, students with

advanced epistemological beliefs had uniformly better content outcomes than students with weak epistemological beliefs. Students with weak epistemology but strong prior astronomy knowledge showed higher content gains after wholly reformed instruction, but only modest gains in epistemology. Students with both weak epistemology and astronomy background had the lowest content outcomes, little affected by epistemological instruction, but showed the largest change in epistemology after epistemological instruction.

Introduction

In the United States a large number of students have their only academic interaction with science during the non-major introductory science course. In astronomy an estimated 200,000 to 250,000 students annually take a non-major introductory science course in the US (Fraknoi, 2001; Mulvey and Nicholson, 2001). Instruction reform designs in large-lecture non-major introductory astronomy courses have predominantly focused on conceptual and content changes in students (Adams, Prather, and Slater, 2005; Green, 2002; Prather *et al.* 2004; Zeilik *et al.*, 1997). Besides content and concept gains, research indicates that student science epistemology is an area of concern (Elby, 2001; Hammer, 1994; Redish, Saul, and Steinberg, 1998; Schommer, 1990).

There are two sources of science epistemology that we can discuss. First is the epistemology of science. This is discussed by Lederman (2004) also as part of Nature of Science. The seven aspects that Lederman (2004) describes as the epistemology of science are:

- Tentative
- Empirically-based
- Subjective
- Involves inference, imagination, and creativity
- Socially and culturally embedded
- Involves knowing the distinction between observation and inference
- Involves knowing the distinction between theory and law

Some of these qualities are complex and subtle to the point that we cannot expect students to have a sophisticated understanding of them. However, there is a second epistemology of science that we can measure in the students, and this may help us to understand how students view the above framework.

Student epistemologies of science can also be seen on a personal epistemological level – how they personally learn and gain knowledge in science. Schommer (1990) identified the personal epistemological beliefs that students have about knowledge. Knowledge is:

- Innate Ability: one is either naturally capable of learning or not
- Simple: knowledge is made up of simple facts that are not part of a larger whole and learning is a matter of memorizing facts
- Certain: knowledge is known and certain and learning is a matter of being told facts from the teacher
- Quick learning: one either learns something right away or not at all

Hammer (1989) added to this through his study of physics students by ideas similar to those noted by Schommer (1990), and also showing – as Schommer (1990) did – that epistemology can affect learning in the physics classroom.

We see that some of these ideas can be mapped onto the framework by Lederman (2004). "Knowledge is certain" can map onto a discussion about the tentative nature of science. "Simple knowledge" maps onto the subjective nature of science in a subtle way.

How can students be taught about both of these epistemologies? Redish, Saul, and Steinberg (1998) found that students in college physics courses had negative score changes on the Maryland Physics Expectation Survey (MPEX). These losses were seen even in the most research-based reform classroom. It is apparent that better instruction in content does not lead to more sophisticated epistemologies of learning in the students. However, Elby (2001) showed that students in a classroom specifically designed to foster epistemological changes had gains on the MPEX and the Epistemological Beliefs Assessment for Physical Sciences (EBAPS). Elby (2001) offers limited evidence from this discussion about the content gain that students made during his epistemological course, but one may argue that since epistemology and learning have been shown to be connected (Lising and Elby, 2005; Schommer, 1990), students are likely to have made some content gain.

Guided by the work of Elby (2001) and White and Frederiksen (1998), we set out in 2005 to design a course that would foster student epistemological changes in the non-major introductory astronomy course at Oregon State University. The design and implementation of this course are described in another manuscript. During the academic year of 2006-2007 we implemented the new instruction in three phases. During the implementation we used three measures, described below, to measure changes in student epistemologies and content gain. This manuscript is an analysis of that data.

The major questions that arise from the development of this project were:

1. Do student epistemologies change after instruction using the redesigned lecture/lab course in a non-major introductory astronomy course?
2. Is there any change in content acquisition when students are instructed using the redesigned non-major introductory astronomy course?

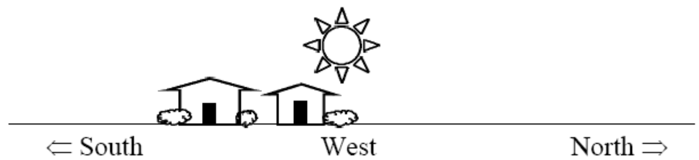
Method

This study uses data collected in the non-major introductory astronomy course during the 2006-2007 academic year at Oregon State University. The academic year at Oregon State University is broken into three eleven week terms. I collected data for each of the three terms. The data that were collected included content outcomes as measured by student exam scores, general prior astronomy content knowledge, pre- and post-instructional epistemological data, and demographic data.

The student exam scores consisted of three multiple-choice exams during the course of the eleven weeks: week five, week eight, and finals week (week eleven). The content covered for each exam was similar from term to term. The week five midterm covered basic physics principles, scientific models and scientific processes, the history of astronomy, and a general overview of measurement, proportional reasoning and measuring the universe. The week eight midterm covered the solar system and the early stages of stellar evolution. The final exam covered the late stages of stellar evolution, galaxies, and cosmology. A cursory review of introductory astronomy texts shows that these topics can be considered “standard” topics in a survey-type astronomy course like the non-major introductory astronomy course at Oregon State University. Since the range of topics was so broad, no single exam could determine the content outcomes of the students. I chose to use the average of all three exams to measure content acquisition. Students that did not take all three exams were excluded from the data analysis. In most cases, the excluded students did not take the final examination. This could be due to withdrawing from the course or to students taking the course in satisfactory/unsatisfactory (S/U) grading. A grade of S in the course requires a C- or better average grade, which could be achieved without attempting the final exam.

General prior astronomy content knowledge was measured using the Astronomy Diagnostic Test version 2 (ADT) developed by Deming *et al.*(2001). The ADT is a set of 21 multiple-choice concept questions that cover a range of astronomy topics. Many questions are observational, while other questions ask physics questions that arise in the astronomy course (see Figure 7). Deming and Hufnagel (2001) used the ADT in a national survey to show that the demographics of non-major introductory astronomy courses are consistent with national trends in college enrollment. I compare their findings with the characteristics of my students enrolled at Oregon State University to show how my students are similar to the national average, suggesting that the reformed instruction method might be useful for a national audience.

9. On about September 22, the Sun sets directly to the west as shown on the diagram below. Where would the Sun appear to set two weeks later?
   A. Farther south B. In the same place C. Farther north



14. Which of the following would make you weigh half as much as you do right now?
   A. Take away half of the Earth's atmosphere.
   B. Double the distance between the Sun and the Earth.
   C. Make the Earth spin half as fast.
   D. Take away half of the Earth's mass.
   E. More than one of the above

Figure 7 Sample questions from the Astronomy Diagnostic Test

The epistemological beliefs of students were measured using the Epistemological Beliefs Assessment for the Physical Sciences (EBAPS). The EBAPS was initially developed by Andrew Elby, John Frederiksen, Christina Schwarz, and Barbara White at the University of California, Berkeley (Elby, 2001). It is a forced-choice Likert-type instrument given to students as pre-instruction and post-instruction tests. On his web site for EBAPS, Elby (2001) explained that EPABS was designed to be used in an algebra-based physics course, and that a version of EBAPS for the purely conceptual physics course would be forthcoming. At the time of this research no such instrument had been published.

The EBAPS consists of 30 questions designed to get at students' beliefs about how scientific knowledge is gained. Elby (2001) attempted to probe student epistemologies “to the extent possible”, while admitting that student course expectations and epistemologies are intertwined. Also there is an attempt to probe the contextual nature of epistemological beliefs that may affect student learning. The EBAPS is structured to measure the following five subscales.

2. Structure of scientific knowledge: Is science based on facts and formulas that are weakly connected, or is science a coherent unified whole?
3. Nature of knowing and learning: Do we learn science by absorbing information, or do we learn by actively engaging the topic, trying to connect science to our everyday lives and making our own meaning?
4. Real-life applicability: Does science apply to the everyday world around us, or is it only useful in laboratories and classrooms? This is about students' understanding of the real-life applicability of science, not their desire to apply it to real life.
5. Evolving knowledge: The extremes are absolutism and relativism. Where do students lie?
6. Source of ability to learn: Is science ability innate or can anyone become a scientist?

The results of the EBAPS consist of scores for each subscale and a single overall score.

The questions on EBAPS appear in three sections. In a survey section, students are presented with a single situation and asked whether they agree or disagree with certain statements. Another section consists of mini discussions between two people, with the students deciding which person they agree with more:

> **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.

The third section presents students with a choice of statements:

> 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).

In all three sections, students must choose a single answer to each question. Other epistemological instruments also were available at the time of my study. The Maryland Physics Expectation Survey (MPEX) developed by the physics education research group at the University of Maryland (Redish, Saul, and Steinberg, 1998) was one prominent alternative to the EBAPS. The MPEX is very useful for measuring physics epistemologies, but it has three problems that would make it less useful for this study. First, the MPEX measures both epistemology and expectations explicitly. As Elby (2001) stated, it is difficult to separate expectations and epistemology completely, but the MPEX makes no attempt to do so. Second, the MPEX is more classroom-specific, as this question from the MPEX shows:

2. All I learn from a derivation or proof of a formula is that the formula obtained is valid and that it is OK to use it in problems.

The nature of the non-introductory astronomy course makes this question confusing to the students, since there are no derivations or proofs of formulas (unlike the usual introductory physics course). Third, the MPEX uses a framework similar to Schommer (1989) and Hofer (2001). MPEX assumes that epistemological beliefs are consistent and articulated. In this study we prefer the framework described by diSessa (1993) and Hammer and Elby (2001) where epistemologies are considered to be fine-grained resources that are activated in context. The EBAPS avoids some of the problems that I would have using MPEX, though it has its own limitations.

Epistemology and expectations are not easily untangled through forced-choice instruments. In his small essay on the idea behind EBAPS, Elby (2001) noted this as one of the limitations of the instrument:

> *Teasing epistemology apart from expectations.* Our validity testing suggests that it's extremely difficult to write items that probe *purely* epistemological beliefs in nearly *all* respondents. For instance, even though
>
> - Someone who doesn't have high natural ability can still learn the material well even in a hard chemistry or physics class.
>
> passed our validity tests, it's likely that some students answered "no" largely because they've never taken a challenging science course.

Thus many of our freshmen may give a naïve answer merely because they have never been in a college physics classroom.

The other two limitations of EBAPS also regard measuring student epistemologies through quantitative instruments. Elby (2001) raised the issue of “teasing beliefs from goals”. A student presented with the proposition, “Given enough time, almost everybody could learn to think more scientifically, if they really wanted to,” may respond negatively because he is motivated to rationalize that the effort would not help him anyway, as an excuse for not putting effort into the course. We would consider that a naïve response, but in other contexts this same student might reveal a more sophisticated view of the same issue.

The final significant limitation of using the EBAPS to measure epistemological changes that I need to address was identified by Elby (2001) as “inferring student sophistication”. Students may answer a question in a sophisticated manner for entirely unsophisticated reasons. This is similar to an observation that Eric Mazur made in the introduction to his Peer Instruction book (1997). Mazur saw that students sometimes could solve what he thought were complex problems, but when he probed their understanding of the content, he found that the students did not understand the underlying concepts of the problem. In probing student epistemologies a student may agree with “Given enough time, almost everybody could learn to think more scientifically, if they really wanted to” , but only, as Elby (2001) proposed, “…believing that learning to think scientifically is no harder than learning to write cursive—it's just a matter of raw practice with no deeper cognitive change”.

These limitations frame the depth of the inferences we can make from the results of the EBAPS, but a rich discussion can still be based on these measurements. While EBAPS scores for individual students may be badly distorted by the test's limitations, we use them only to compare various populations of students, and to gauge epistemological changes by comparing EBAPS results on pre- and post-testing of various groups. Individual changes in certain students would best be served by interviews, which would have required more resources than were available for this study.

Students

In order to study the three terms as individual experimental groups it is important to know whether these groups are equivalent or not. We used statistical analysis, reported in the discussion below, and visual interpretation to determine what may be causing those differences. Each figure in this section shows the demographics of the students enrolled in the non-major introductory astronomy course at Oregon State University for each of the three terms that this study took place. The choice of demographic questions was constrained by my choice of instrument. The Astronomy Diagnostic Test (ADT) was used to gather demographic data (Hufnagel, 2002). The ADT was designed to be used with fill and scan sheets (Scantrons), which required the designers of the ADT to make choices of how to group certain demographic questions (Hufnagel, 2002). These groupings are discussed in my analysis. There are eleven demographic questions on the ADT:

- In general, how confident are you that your answers to this survey are correct?
- What is your college major (or current area of interest if undecided)?
- What was the last math class you completed prior to taking this course?
- What is your age?

- Which best describes your home community (where you attended high school)?
- What is your gender?
- Which best describes your ethnic background?
- Which best describes your ethnic background?
- How good at math are you?
- How good at science are you?
- Which best describes the level of difficulty you expect/experienced from this course?
- How many astronomy courses at the college level have you taken?

The question "Which best describes your ethnic background?" is used twice on the ADT for ease of use with scantron sheets. I chose to use five demographic questions from the ADT: major, gender, age, "How good at science are you?", and "How good at math are you?".

A discussion of each demographic measure will accompany the figures.

Major.

Figure 8 shows the academic major of students entering the astronomy course in each term. Immediately we see that this course is dominated by non-science and non-education majors. However, the distribution of majors is independent of the term selected $\chi^2(8, N = 476) = 12.8$, $p = 0.12$. Being a 100-level non-major science course, the lack of science majors is not surprising. The course satisfies a baccalaureate core requirement for Oregon State University, where all students are required to have at least one lab science course. The science and engineering majors satisfy this with their introductory physics courses. Therefore, I expect that the science and engineering majors that enroll in the course are taking it due to an interest in astronomy. That is not to say that these students would not benefit from epistemologically-based instruction.

We also see that there is an increase in education majors over the course of the year from 4% in FALL 2006 to 10% in SPRING 2007. It is unclear whether this is truly a trend in enrollment, or an effect of more education majors being willing to participate in the study in SPRING 2007.

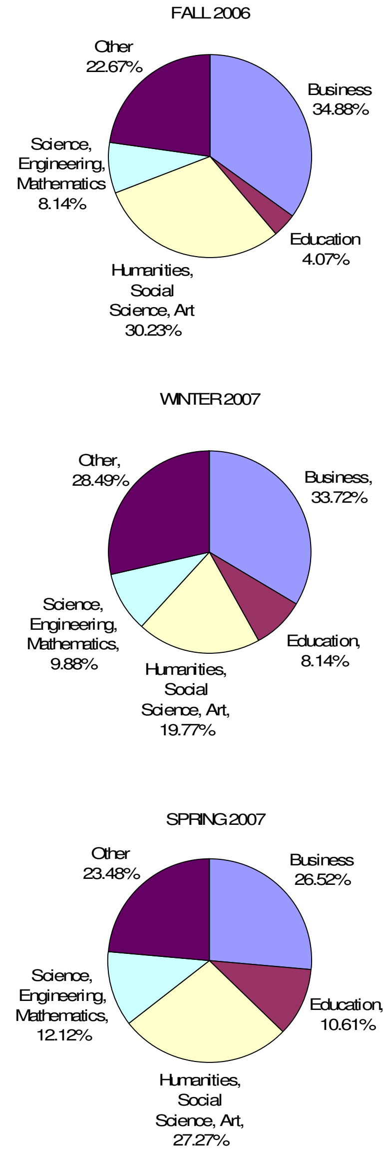


Figure 8 Students by term and major

Gender and Age.

Figure 9 shows two distributions; the set on the left show the distribution of students by age for each term, and the set on the right show the distribution by gender.

On the ADT students are allowed to decline to answer the gender or age questions. In the gender category, students in FALL 2006 and WINTER 2007 chose "decline to answer" less than 1% of the time, while in SPRING 2007, a surprising 16% of the students declined to answer. In reporting student demographics, I assumed that those students were distributed proportionately to the respondents. This is shown in Figure 9. We see in the Gender category that males dominate in FALL 2006 and SPRING 2007, but females are equally represented in WINTER 2007. However, the distribution of gender is not dependent on the term $\chi^2(2, N = 448) = 1$, $p = 0.67$.

In the Age category, the distribution is dependent on the term $\chi^2(10, N = 476) = 30.1$, $p = 0.12$. Calculations using Scheffé intervals indicated that SPRING 2007 was significantly different from FALL 2006, $F(2,473) = 3.67$, $p < 0.005$, and WINTER 2007, $F(2,473) = 7.67$, $p < 0.05$. We see that SPRING 2007 is characterized by more students age 24 or older compared to the other two terms. In this case, I felt it significant to report the number of students that declined to answer in SPRING 2007 since it is much higher than the other two terms. It is not apparent why there is such an age difference in the three terms, but in discussions with Physics Department Chair Henri Jansen (personal communication, March, 2008), we conjectured that every spring there is an increase of seniors in the course due to graduation requirements. This would change the average age of the class. Since other demographic variables were consistently the same, I decided that the comparisons of the groups could proceed.

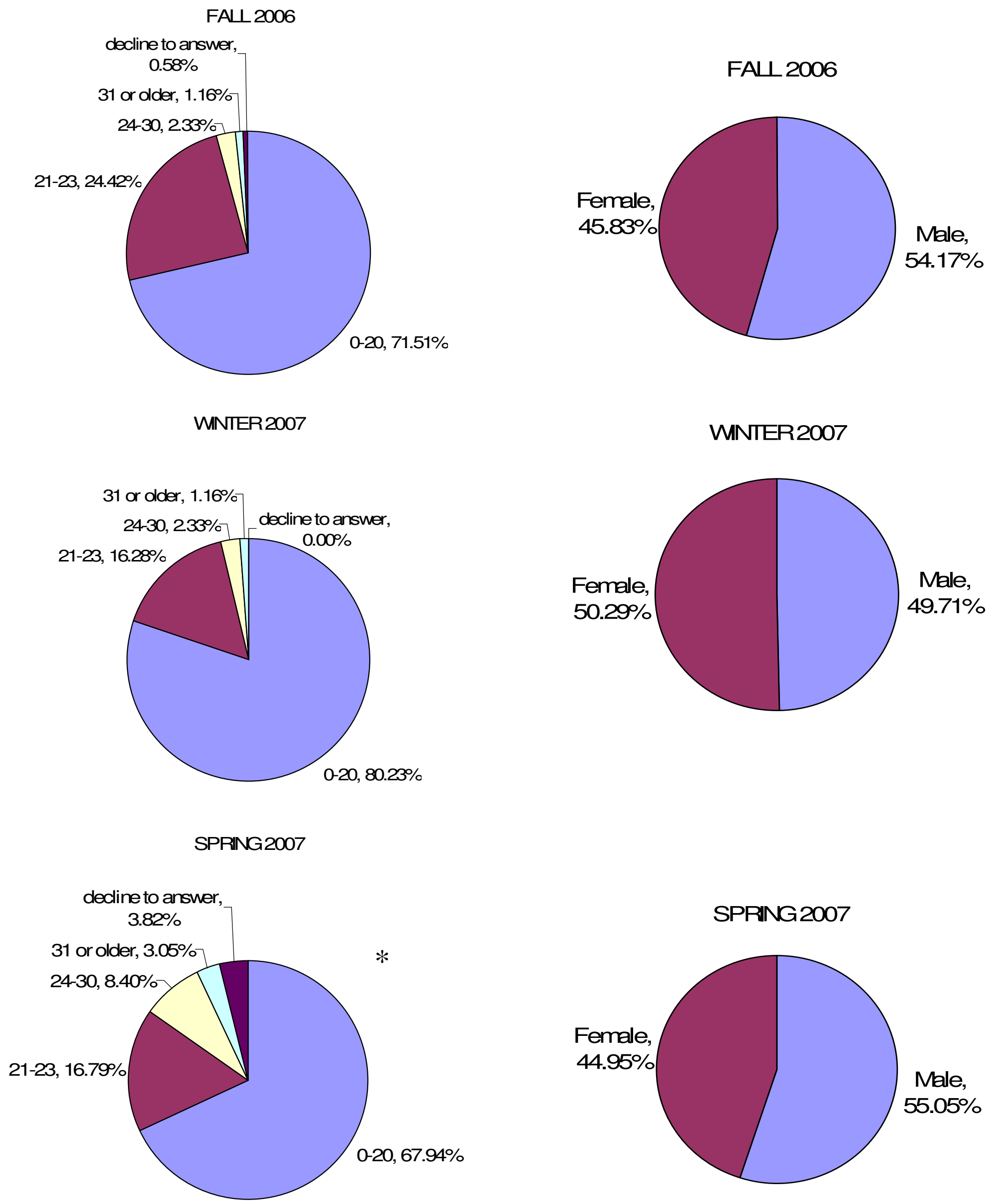


Figure 9 Student distribution by age and gender.
*, $p < 0.05$

"How good at math are you?"

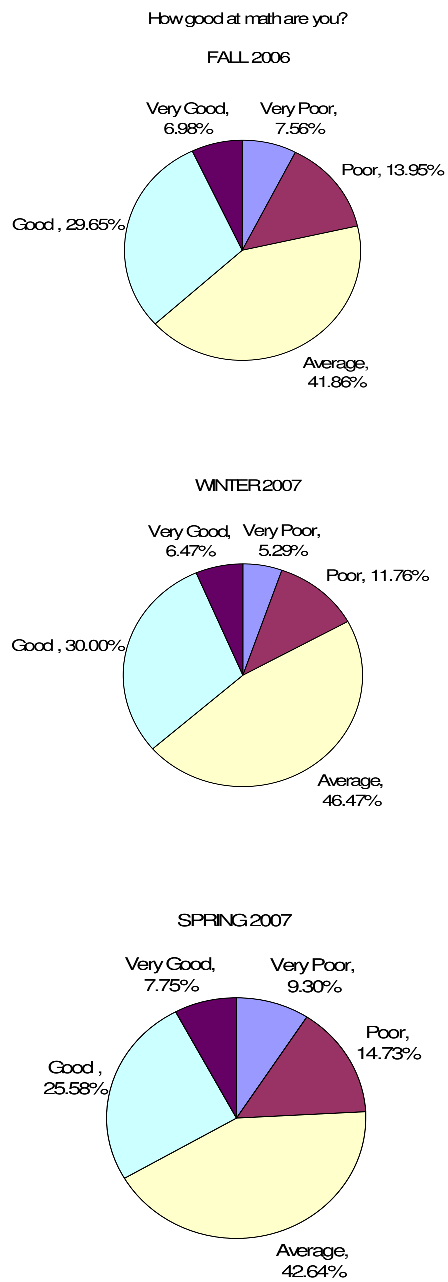
A specific question that students were asked on the ADT is "How good at math are you?" The five possible responses range from "very poor" to "very good". The distribution of responses to this question by term can be seen in Figure 10. We see that for each term the distribution is nearly identical (χ^2, $p > 0.05$). The class is dominated by a group that feels they are average at math and another population that feels they are good at math.

"How good at science are you?"

Similar to the previous question, "How good at science are you?" has a similar distribution from term to term (see Figure 11). This distribution is dominated by those that think they are average at science. Also very few of the students consider themselves "very poor" or "very good". Neither the figures nor χ^2 reveal any significant difference between terms for this demographic measure ($p > 0.05$).

I have shown the demographics of major, gender, age, "How good at math are you?", and "How good at science are you?" In all cases except age, there is no large distinguishing difference between each term. This age difference appears to come from the large increase in the percentage of older students and decline-to-answer students in SPRING 2007. In spite of this difference, the bulk of evidence indicates that we can treat these three terms as equivalent groups for the basis of our discussion. This will become important when comparing the results of varying instruction from term to term. In order to discuss the epistemological and content-acquisition changes, I will first show the pre-instruction epistemological scores for our baseline study in

Spring 2004 and the three terms of this study (FALL 2006, WINTER 2007, and SPRING 2007)



.

Figure 10 Distribution of student responses to "How good at math are you?"

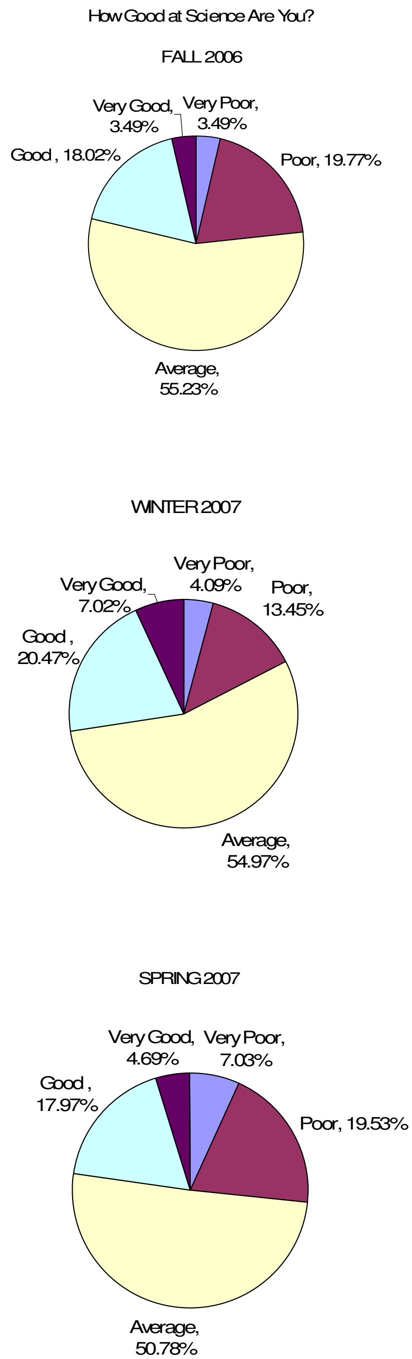


Figure 11 Distribution of student responses to "How good at science are you?

Pre-instruction Epistemological Scores

Before and after instruction in Spring 2004, we gathered data using the EBAPS. The pre-instruction scores can be compared to the pre-instruction scores for the terms in this study. I assumed that since I saw little variance in the other demographic data from term to term in 2006-2007, overall demographics in Spring 2004 would be similar. However, the comparison of pre-instruction EBAPS among the four terms does reveal one significant anomaly – in FALL 2006 (Figure 12). From this distribution we see that there is a substantial group of students in FALL 2006 whose epistemologies are less sophisticated than students enrolled in any other term. Students with pre-instruction EBAPS scores were significantly different from term to term, $F(2, 148) = 15.2$, $p < 0.00001$. Further analysis showed that FALL 2006 was significantly different from WINTER 2007, $F(2, 148) = 10.7$, $p < 0.00001$, and SPRING 2007, $F(2, 148) = 10.6$, $p < 0.00001$. However, WINTER 2007 and SPRING 2007 are not significantly different, $F(2, 148) = 0.023$, $p = 0.37$. This aspect of the students' backgrounds is so crucial that we consider it as an explicit variable in the analysis of instructional results.

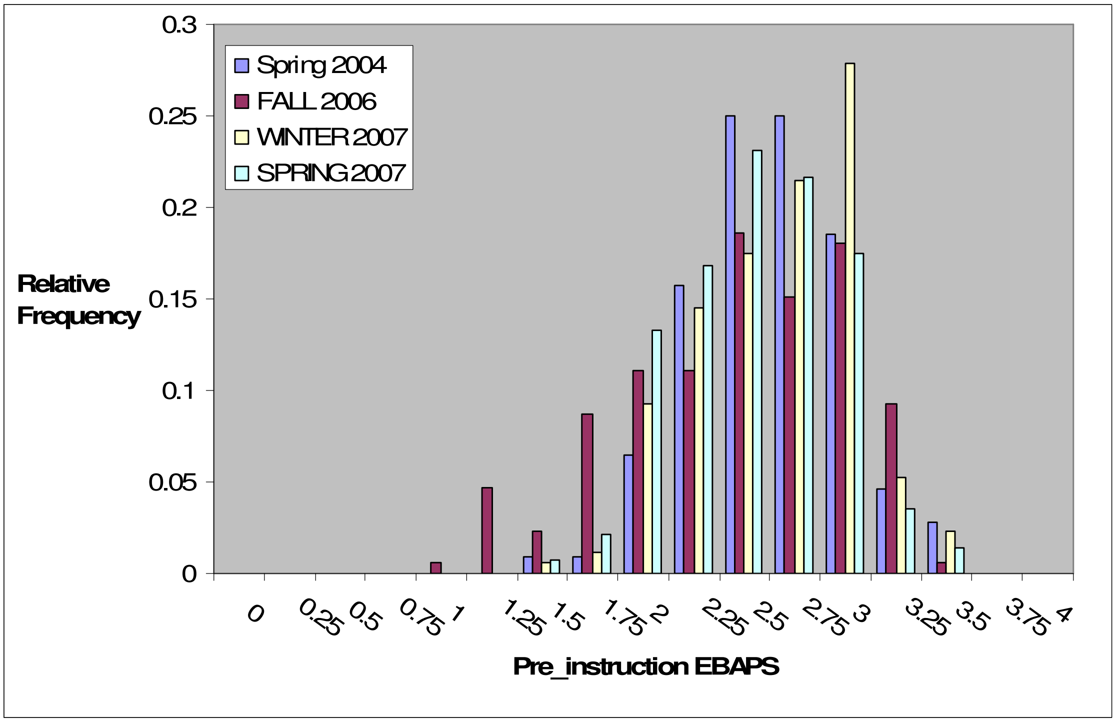


Figure 12 Pre-instruction EBAPS scores by term. Note the large population of students in FALL 2006 with low scores.

The other pre-instruction variable we measured, except for Spring 2004, is the general astronomy content knowledge of students as quantified by the ADT. As described above there are 21 general content knowledge questions on the ADT. Figure 13 shows the distribution of scores on the ADT for each of the three terms in the 2006-2007 study. First, no student in any term correctly answered all questions. In SPRING 2007, there was one student that scored a 0 on the ADT. It also appears from this distribution that SPRING 2007 had more students with lower ADT scores. ANOVA indicated that the three terms were significantly different, $F(2, 484) = 6.91$, $p = 0.0013$ Further analysis indicated that SPRING 2007 was significantly different from FALL 2006, $F(2, 484) = 4.77$, $p < 0.001$, and WINTER 2007, $F(2, 484) = 7.73$, $p < 0.001$.

The lower overall scores for the ADT in SPRING 2007 are important for our subsequent discussion about content gain.

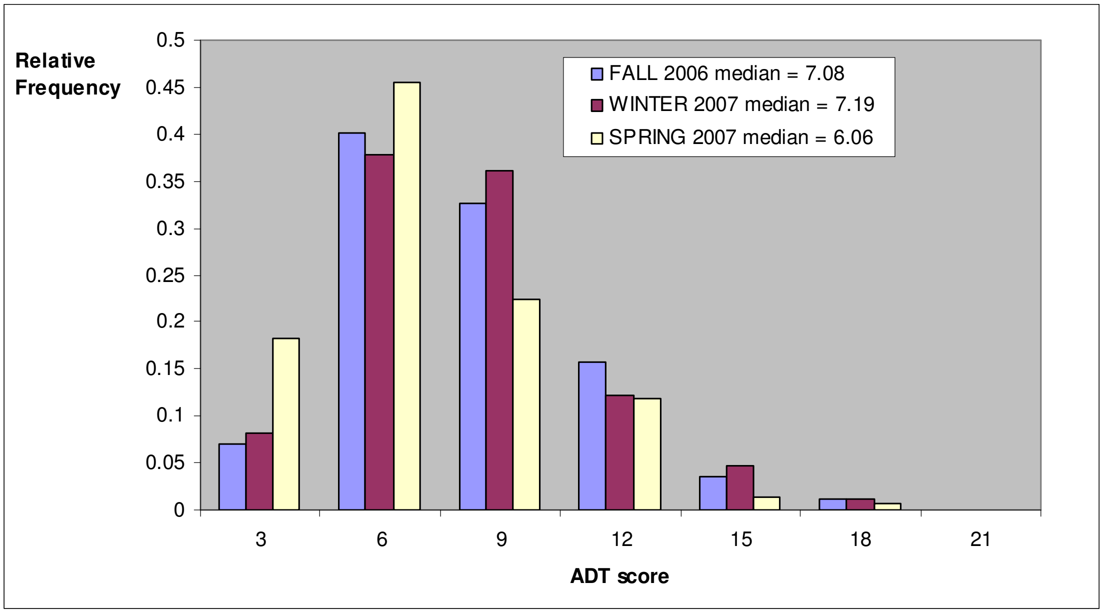


Figure 13 Distributions of ADT scores for students in the 2006-2007 academic year by term. We see that there are a large percentage of students in SPRING 2007 with low ADT scores.

Instruction

A complete discussion of the instruction can be found in the previous manuscript. The overall design of this course uses scaffolded instruction to foster changes in student epistemologies of science through changing their epistemologies of how they learn science. The instructional changes were made over the course of the 2006-2007 academic year. At Oregon State University the academic year is three 11-week terms with 10 weeks of class meetings and one finals week. The astronomy course under discussion is a stand-alone 11-week course, so each term we had a new experimental group. We introduced the instructional changes in three phases. The first phase, in the fall term of 2006 (FALL) introduced only redesigned labs. The second phase, in the winter term of 2007 (WINTER), introduced more varied instruction into the lecture class and maintained the labs as they were. The final phase, introduced during spring term of 2007 (SPRING), included a pre-lab component that structured the discussion for each week's meetings. Table 2 shows a timeline of the instructional changes.

The previous sections were devoted to a discussion of the research groups in the 2006-2007 study. I have confirmed that we can treat each term as equivalent groups for comparison. The exception is the age group. SPRING 2007 has a significantly different distribution of students older than 24. The large group of students with low EBAPS scores in FALL 2006 was considered an important group to discuss. This was also the case with the significantly lower ADT scores in SPRING 2007. These two variables are important enough that they will both be highlighted in the following results section of this manuscript.

Table 5 ***Timeline of instructional changes*****.**

Term	Instructional Changes	Instructional content
Fall term 2006 (FALL)	Introduced redesigned labs	Lecture instruction using only Peer Instruction for entire term. Lab instruction using scaffolded instruction based on the inquiry cycle of White and Fredericksen (1998)
Winter Term 2007	Maintained labs from	Lecture instruction using Peer

(WINTER)	FALL, introduced modified instructional practices in lecture	Instruction, concept maps, tutorials, advanced organizers, demonstrations. Maintained lecture pace regardless of lab.
Spring Term 2007 (SPRING)	Introduced pre-lab that coupled instruction from WINTER and labs from FALL.	Pre-labs designed as part of inquiry cycle to prepare for lab. Pre-labs used as an instructional tool in lecture class along with other lecture methods.

Results

Epistemological Changes

The major focus of this study is to gauge the effect, if any, of an instruction model intended to foster changes in student epistemologies. Our first look at the data views the changes in pre- to post-instruction EBAPS scores from term to term. Since instructional modifications were introduced progressively, we compare chronologically successive terms.

Figure 14 shows a scatter plot of the post-instruction EBAPS results versus the pre-instruction EBAPS results for each of the four terms – Spring 2004, FALL 2006, WINTER 2007, and SPRING 2007. Improvement is indicated by displacement above the diagonal. It is apparent that the greatest improvements (upper left region of figure) were achieved by the group of FALL 2006 students earlier identified by their unusually naïve epistemologies.

Instruction in every term served to make students' epistemologies more similar to each other. This is most apparent in FALL 2006 where the most naïve group showed the greatest gains. This effect is also present in the other terms as can be seen from the linear fits in Figure 14. In each case the slope of the trend line is less than 1 demonstrating a homogenizing result.

To further view the effects of instruction on epistemology, we look at the same data using histograms of the change score. Change score is a number defined by us as the post-instruction EBAPS minus the pre-instruction EBAPS (Figure 15). Once again the previously-identified group with very naïve epistemology in FALL 2006 shows the largest change score, represented by the tail at positive values.

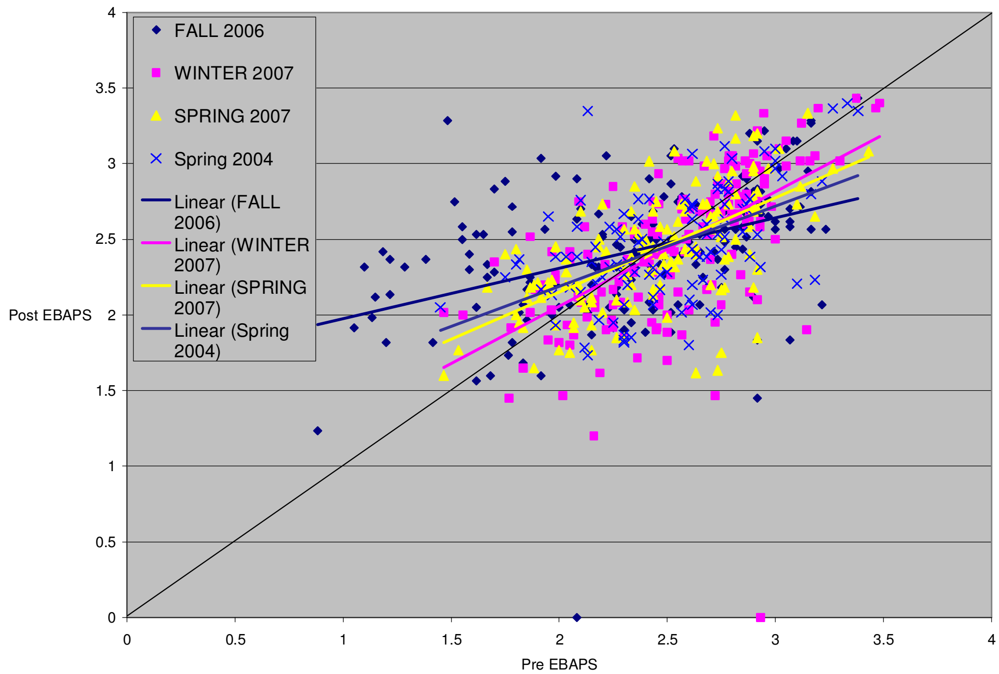


Figure 14 Scatter plot showing post-instruction versus pre-instruction scores on EBAPS measure of epistemology.

The other striking feature from our data (Figure 15) is the large number of students with negative change scores. At first glance it seems discouraging that students concluded the course with less sophisticated epistemologies than they began with. However, this phenomenon is well known from studies of physics instruction (Redish *et al.* 1998).

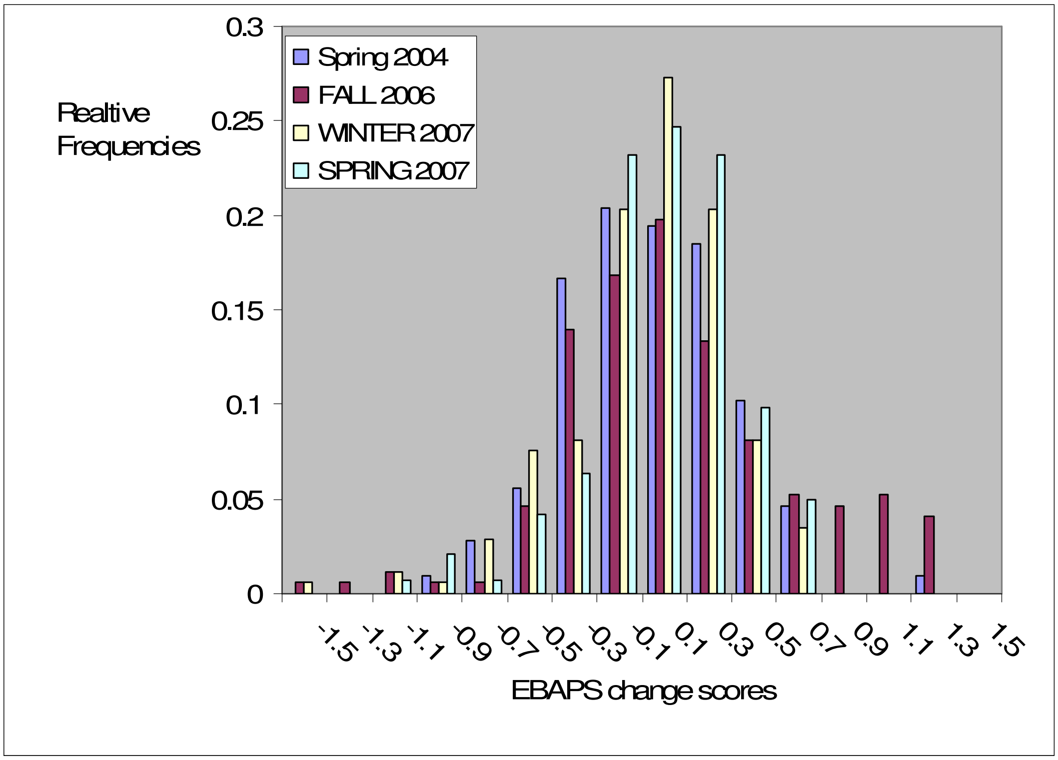


Figure 15 Distributions of change scores for all four terms in this study.

Inspecting our results in Figure 15, we see a steady reduction from term to term in the group of students with negative change scores. To view this feature we plot in Figure 16 the percentage of students with change score less than -0.2. We see that we have two marked decreases in the percentage of students with substantial negative change scores. The first such decrease is from Spring 2004 to FALL 2006. This change coincides with the introduction of the redesigned labs in FALL 2006, and seen in the previous manuscript.

The second substantial decrease comes between WINTER 2007 and SPRING 2007. Fewer than 15% of the students have significant negative change scores in SPRING 2007. We connect this substantial decrease in negative change scores to the introduction of a connected instruction cycle previously detailed in the prior manuscript.

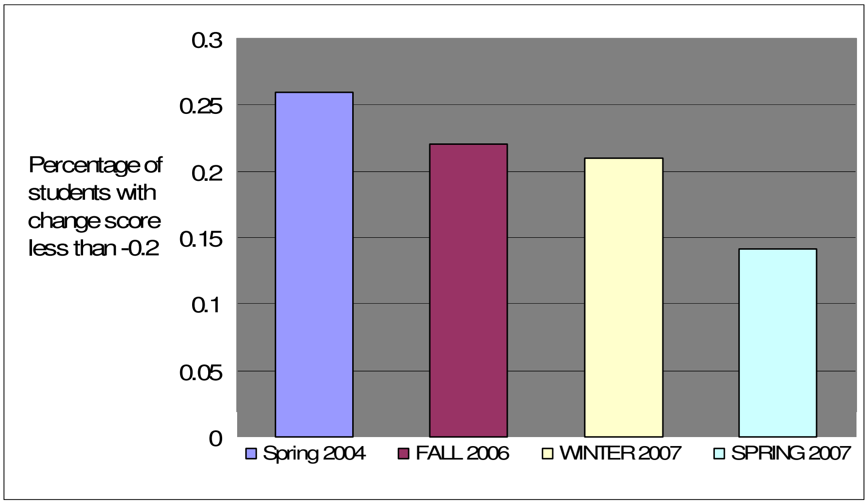


Figure 16 Fraction of students with substantial loss in epistemological development as indicated by losses greater than 0.2 in EBAPS scores in each term.

The histogram in figure 15 does not reveal obvious trends in the central parts of the distributions. An alternate presentation of the same data (Figure 17) confirms our impression that the greatest differences among the groups are in the outliers. In light of this observation I am reluctant to assign significance to the mean values of the outcomes.

In the previous analysis we have seen that my reformed instruction model is accompanied by positive epistemological changes. This is most evident for the students with the most naïve epistemological background. Although a few of my students experienced epistemological degradation, nearly twice as many students were set back by the previous traditional course (14% vs. 26%). As Redish, Saul, and Steinberg (1998) showed, physics instruction that is tailored only to content change will produce degradation in student epistemologies, even when successful in content knowledge retention. In the next section we will see what effects my instruction model had on content knowledge retention.

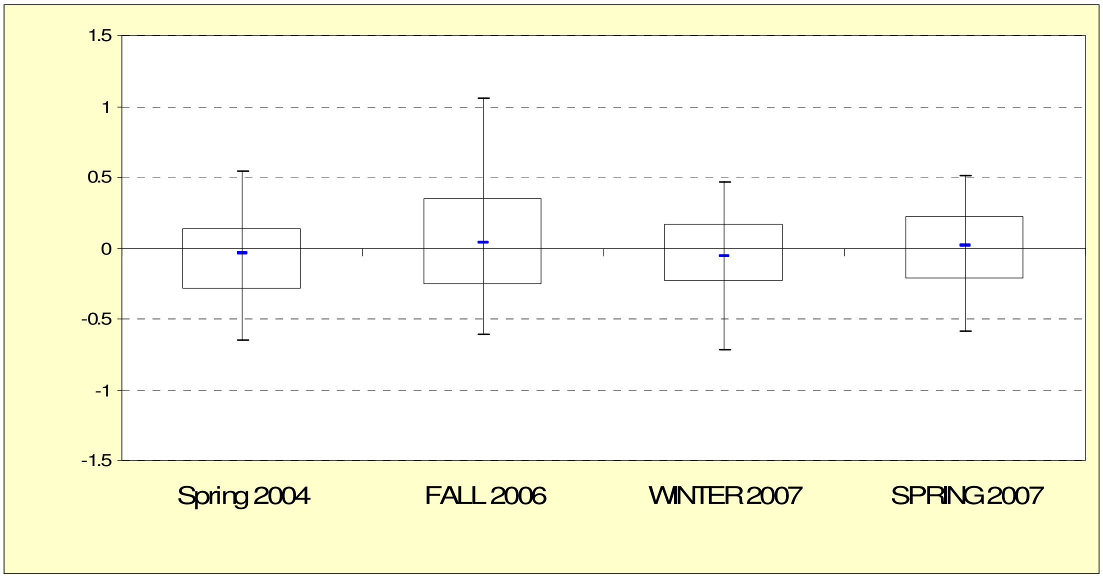


Figure 17 Distributions of changes in EBAPS scores, by term. Boxes enclose second and third quartiles, terminated lines show range for each group. Medians are points near centers of boxes.

Content Knowledge Retention

Besides the epistemological results seen in pre-post changes in student EBAPS scores, another concern was the content knowledge of the students. If labs and lectures were redesigned to foster epistemological changes in the students, would I sacrifice the astronomy content that students were learning? Figure 20 shows a box plot of the overall exam average by term. As explained in data collection this is the average of all three exam scores for a student in that term. Only those students that attempted all three exams are represented in these plots. WINTER 2007 shows two outliers in the data set. These outliers skew the distribution to lower scores. However, the distributions appear to be normal. A glance at the box plots indicates that there is probably no difference in the means of the distributions. Analysis indicates that there is no significant difference in the exam averages from term to term, $F(2, 484) = 2.85$, $p = 0.06$. What this indicates is that during my instruction for epistemological changes there was no change in content understanding.

The importance of the equivalence in the exam averages over the year becomes more evident when remembering that the SPRING 2007 students entered the course with a significantly lower median ADT score than the other two terms. The SPRING 2007 instruction model helped those students learn and understand the content at a level similar to those students entering with much more knowledge of general astronomy content. The lecture-lab connection helped these students. Figure 18 shows a scatter plot of exam average versus ADT. We note that the exam scores in SPRING 2007 are slightly higher than the other two terms for students with the same content background (see Figure 19). Also their distribution appears more uniform than the other two, indicating that the disadvantaged students were helped most by this instruction.

What does this mean for instruction? I have shown that when we attend to the epistemologies of the students in the classroom, we can make gains in their epistemology of science and their epistemology of learning science, especially for the students with the most naïve epistemologies entering the course. What we see from Figure 20 is that when we adopt this style of instruction in our classroom it does not need to be at the cost of content knowledge. Indeed, our most effective instructional plan actually helps the weaker students acquire more content knowledge.

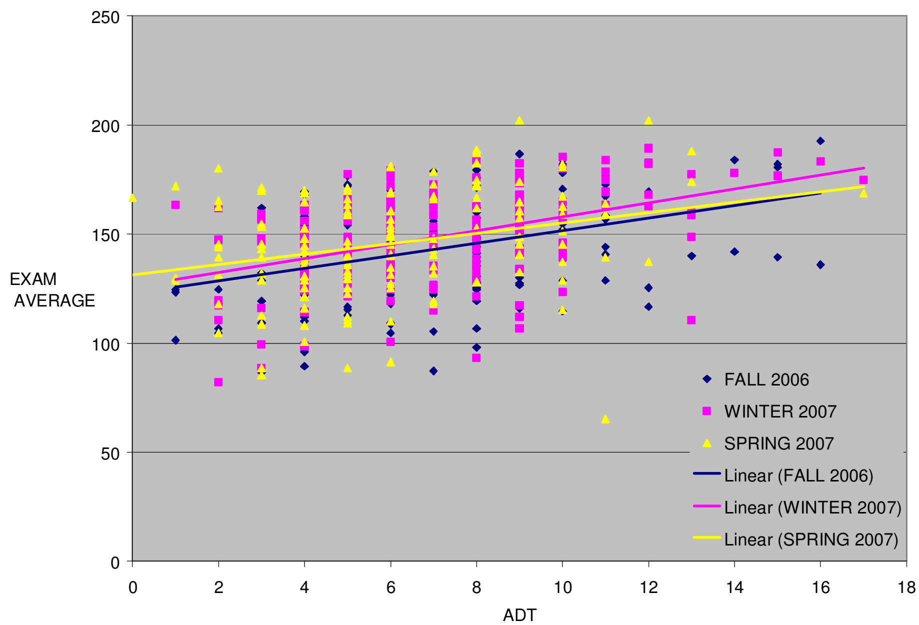


Figure 18 Course content success measured by exam averages as function of content preparation measured by ADT. Solid lines are linear fits to each term's data

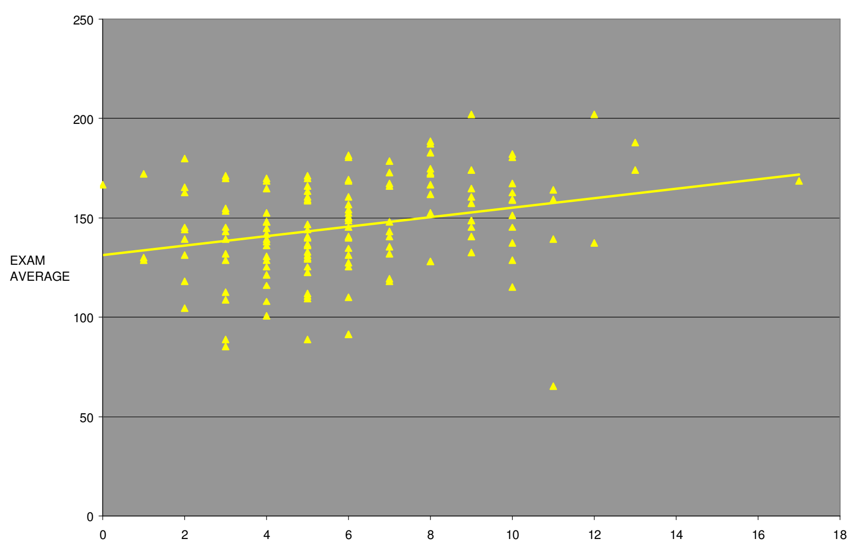


Figure 19 Exam average versus ADT score for SPRING 2007 separated to show the evenness of the data. This indicates that all students had similar content knowledge at the end of the course.

A concern in this analysis is that by looking at the overall exam average of the whole class, we cannot gauge what has happened to those students who may have begun with pre-instruction epistemologies that were more naïve or more sophisticated. In the next section I will explain how, with the help of Larry Flick and Phil Siemens, I separated the data into finer groups for discussion.

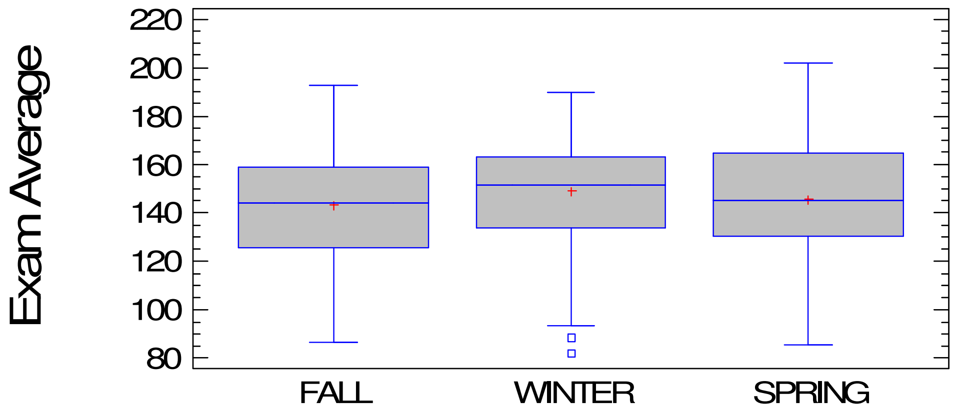


Figure 20 Box plot of overall exam average by term.

During the course of data analysis I decided that splitting each term into thirds based upon ADT scores, exam averages, and pre-instruction EBAPS score would give a better indication of the effect of this instruction on overall content knowledge as quantified by the exam average. In the histogram of exam averages by term (see Figure 20), we see that there are no apparent differences in the exam averages from FALL 2006 to WINTER 2007 to SPRING 2007. Each group was separated into the top third (HI) and bottom third (LO). I did not look at the middle third for this analysis. Table 6 shows the naming scheme that is used for reporting the findings for these groups. I will consistently use this naming scheme in figures and text.

Table 6 Naming Scheme for Groupings

Measure	Group	Name Reported
.Astronomy Diagnostic Test (ADT)	Upper Third	ADTHI
	Lower Third	ADTLO
Pre-Epistemological Beliefs Assessment for Physical Sciences (Pre-EBAPS)	Upper Third	PreEBAPSHI
	Lower Third	PreEBAPSLO
Exam Average	Upper Third	EXAMHI
	Lower Third	EXAMLO

After the groups were defined, their demographics for each term were compared based upon the ADT and the pre-EBAPS. I looked at the means for the pre-instruction EBAPS scores for each ADT group. Conversely, I looked at the ADT means for each of the pre-EBAPS groups. These new demographics were introduced in order to allow us to compare the exam averages and the pre-post change in the EBAPS between each term and each group.

Not surprisingly there is a correlation between performance on the pre-EBAPS and the ADT. In every term the ADTHI group had a significantly higher mean pre-EBAPS score than the ADTLO group in the same term (see Table 7). Also, those students in the PreEBAPSHI group had higher mean ADT scores than those students in the preEBAPSLO group (see Table 7).

As expected from our discussion of the demographics, we see that the ADT means in SPRING 2007 are much lower than the ADT means for the other two terms (see Table 7).

Table 7 Means for preEBAPS and ADT by term split by ADT HI/LO or preEBAPS HI/LO groups respectively.

Term	Pre- EBAPS Means		ADT Means	
	ADTHI	ADTLO	preEBAPSHI	preEBAPSLO
FALL 2006	2.59	2.24*	8.18	6.24*
WINTER 2007	2.64	2.45*	8.08	6.29*
SPRING 2007	2.61	2.30*	6.98	5.26*

* $p < 0.05$ between HI/LO groups

After the previous analysis was completed I defined categories of ADT versus pre-instruction EBAPS (Figure 21). This allows me to display data for each of four groups: ADTHI/preEBAPSHI, ADTHI/preEBAPSLO, ADTLO/preEBAPSHI, and ADTLO/preEBAPSLO (see Figure 21, Figure 25, Figure 26, and Figure 27). The top left box encompasses the ADTHI/preEBAPSHI group, i.e. those students in the top third of both. The lower right box encompasses the ADTLO/preEBAPSLO groups, i.e. the lower third of both. The other two boxes are students in the lower third of one group but the upper third of the other. In this section I will highlight the findings concerning exam average and pre-post EBAPS changes.

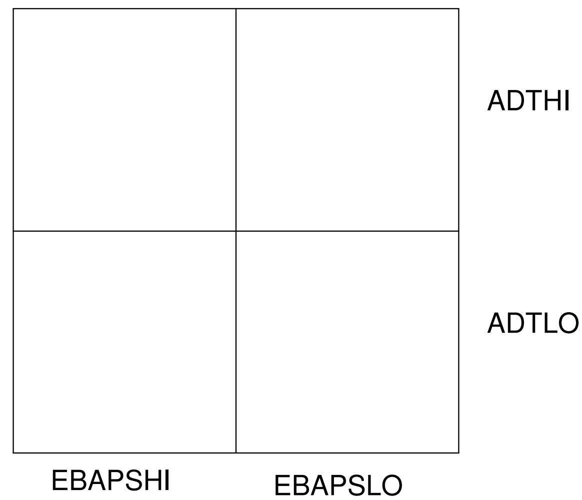


Figure 21 Categories defined by ADT versus preEBAPS

Exam averages for each group are given in Figure 22, Figure 23, and Figure 24. The first item to look at is overall trends in the data. Unsurprisingly the ADTHI/preEBAPSHI group in each term has the highest median exam average where as explained above the exam average is the average of all three exams that students were given through the term. Thus the median exam average would be the median of the average score for the three exams the students were given. This indicates that students entering with more sophisticated epistemologies and more general astronomy knowledge were already prepared to succeed in the course, and I facilitated their learning.

Table 8 shows the median scores for the exam averages for the ADTHI/preEBAPSHI students. There is not a significant difference between the scores for these students $F(2, 79) = 0.75$, $p = 0.47$. There is no effect of instruction on the very highest group.

Table 8 Median score for student exam averages for the ADTHI/preEBAPSHI students for FALL 2006, WINTER 2007, and SPRING 2007. We see that the averages are similar for this group.

Term	ADTHI/preEBAPSHI Score
FALL 2006	170
WINTER 2007	175
SPRING 2007	165

Another unsurprising outcome is that the ADTLO/preEBAPSLO group has the lowest median exam average consistently from term to term. These students are much less prepared to succeed in the college science course than their ADTHI/preEBAPSHI counterparts. The pace of an eleven week course in college may be too fast for these students. When the end of the course comes, these students may just be beginning to internalize and negotiate the epistemological discussions that we have had, and this may be interfering with their ability to understand the content that we have also been working on through the course.

We can see in the box plots (see Figure 22, Figure 23, and Figure 24) that there is only a very slight upward trend in the exam average from FALL2006 to WINTER 2007 to SPRING 2007, though analysis indicates no significant difference among the three medians, $F(2, 70) = 0.56$, $p = 0.57$. The median is 130 in FALL2006 is 130 again in WINTER 2007 and is 135 in SPRING 2007. This indicates that as the instruction implementation changed through the year, there was no clear effect on the students in the ADTLO/preEBAPSLO group. There was no effect of instruction on the very lowest group just as there was not effect on the very highest group.

The data for the remaining groups, ADTLO/preEBAPSHI and ADTHI/preEBAPSLO (see Figure 22, Figure 23, and Figure 24), but are more complicated than those we have already considered. We note immediately that median exam averages for these two groups mostly lie between the ADTHI/preEBAPSHI group and the ADTLO/preEBAPSLO group. The exception is the ADTHI/preEBAPSLO group in SPRING 2007. I will discuss both groups next

Figure 22, Figure 23, and Figure 24 indicate that the ADTLO/preEBAPSHI group had lower median exam average in FALL 2006 and SPRING 2007 than in WINTER 2007. Analysis of the ADTLO/preEBAPSHI group did not indicate a significant different from term to term, $F(2,39) = 0.69$, $p = 0.51$. However, the ADTLO/preEBAPSHI group, like the ADTLO/preEBAPSLO group overall had the lowest of the exam averages.

The final group in our discussion is the ADTHI/preEBAPSLO group. The ADTHI/preEBAPSLO group consisted of those students with higher general content knowledge and more naïve epistemologies. Figure 22, Figure 23, and Figure 24 show that this group has a consistently higher median exam average than the ADTLO/preEBAPSHI group in the same term, but we see that the median exam average in SPRING 2007 for the ADTHI/preEBAPSLO group is different than the median exam average of the equivalent group in FALL 2006, $F(2, 35) = 3.45$, $p = 0.044$ and WINTER 2007, $F(2, 35) = 3.61$, $p = 0.044$, and it is equal to the median exam average for the ADTHI/preEBAPSHI group. However, the ADTHI/preEBAPSLO group in FALL 2006 and WINTER 2007 had lower median exam averages than the ADTHI/preEBAPSHI group in each term respectively. This indicates that students in the ADTHI/preEBAPSLO had an increase in content knowledge greater than the equivalent group in the other two terms. The connected instruction helped this group more than the other groups in SPRING 2007.

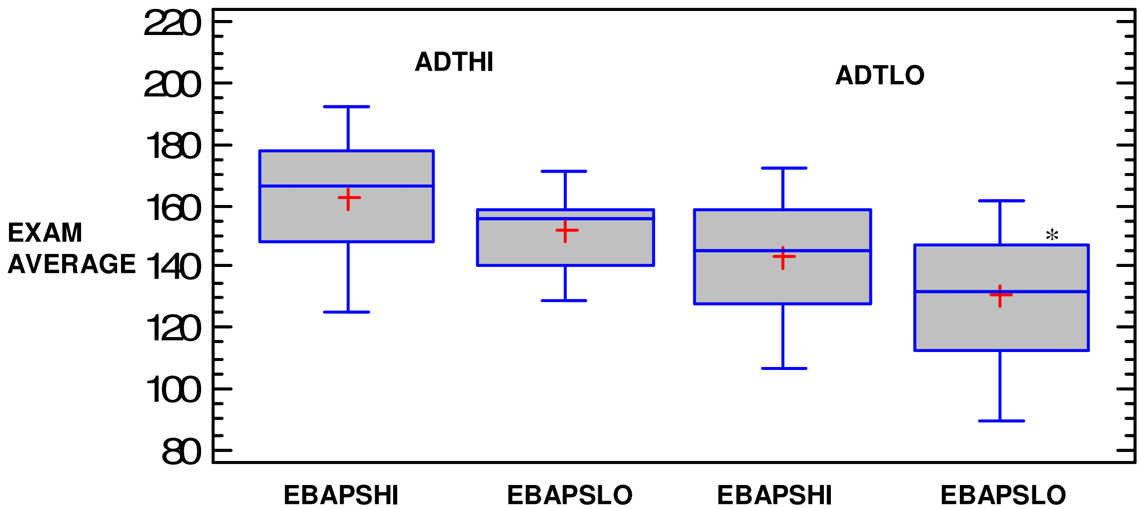


Figure 22 Exam average for FALL 2006 broken into the four ADT/preEBAPS groups. There is a general trend downward from ADTHI/preEBAPSHI to ADTLO/preEBAPSLO. * p < 0.01

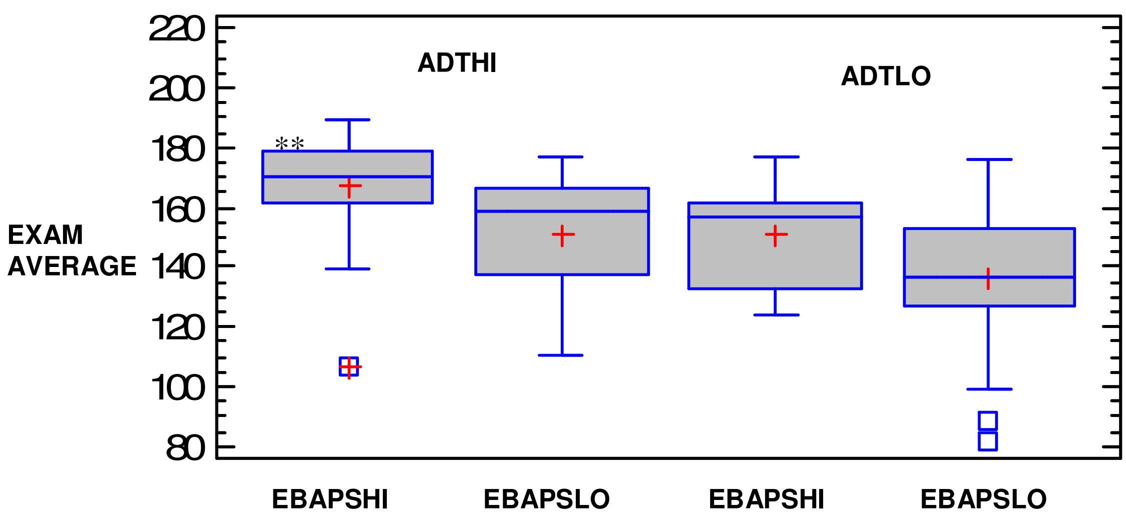


Figure 23 Exam average for WINTER 2007 broken into the four ADT/preEBAPS groups. There is a general trend downward from ADTHI/preEBAPSHI to ADTLO/preEBAPSLO. Overall scores are not different from FALL 2006 with the exception of the ADTLO/preEBAPSHI group. ** Significantly higher than LO/LO group

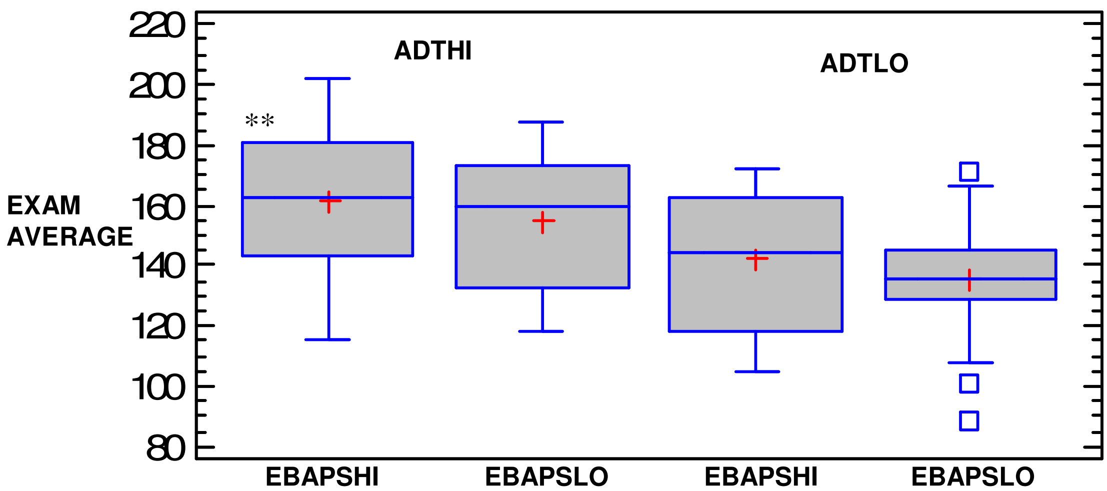


Figure 24 Exam average for SPRING 2007 broken into the four ADT/preEBAPS groups. There is a general trend downward from ADTHI/preEBAPSHI to ADTLO/preEBAPSLO. Overall scores are not different from FALL 2006 or WINTER 2007with the exception of the ADTLO/preEBAPSHI and ADTHI/preEBAPSLO groups. ** significantly higher than LO/LO group

A question arises when looking at this final group: was there a similar change in the pre- and post-instruction EBAPS results in the ADTHI/preEBAPSLO group in SPRING 2007 compared to the other groups in that term or compared to similar groups in during the year? Did this group have a higher EBAPS change score than the other groups? This is the second analysis I introduced previously. I chose to plot box plots for the pre- post change in each group similar to the median exam average for each group. I did this across all three terms, FALL 2006, WINTER 2007, and SPRING 2007(see Figure 25, Figure 26, and Figure 27). Immediately we see what appear to be very high median changes in the pre- to post-instruction EBAPS results for the both ADTHI/preEBAPSLO and ADTLO/preEBAPSLO groups in FALL 2006. These large values for the median changes can be traced back to our previous discussion about FALL 2006 where we saw that there was a significant population of students with well below median pre-instruction EBAPS scores. Since I drew from this population to create my groups the high change score is not surprising. This confirms that epistemological changes can be dramatic for the most disadvantaged students.

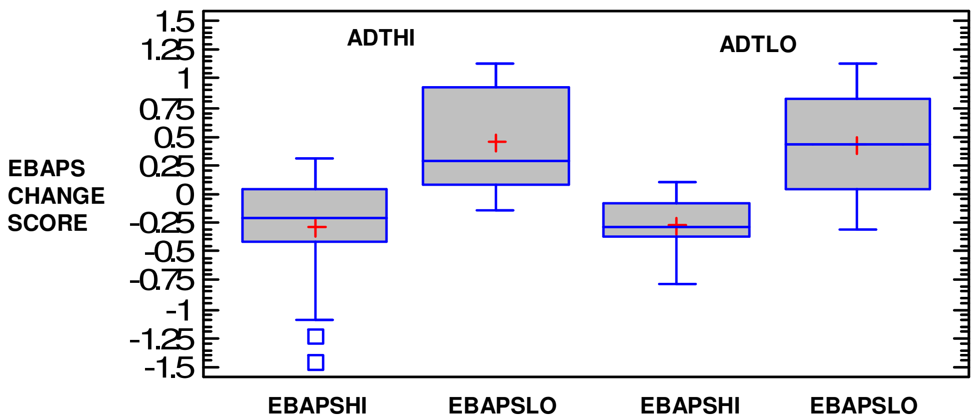


Figure 25 EBAPS change scores for the four groups in FALL 2006. The large change score for both preEBAPSLO groups is due to the population that we drew from to create the groups.

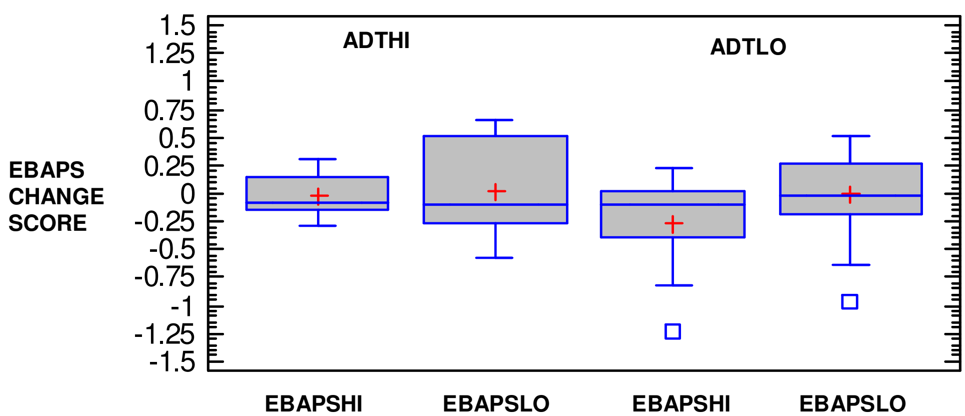


Figure 26 EBAPS change scores for the four groups in WINTER 2007

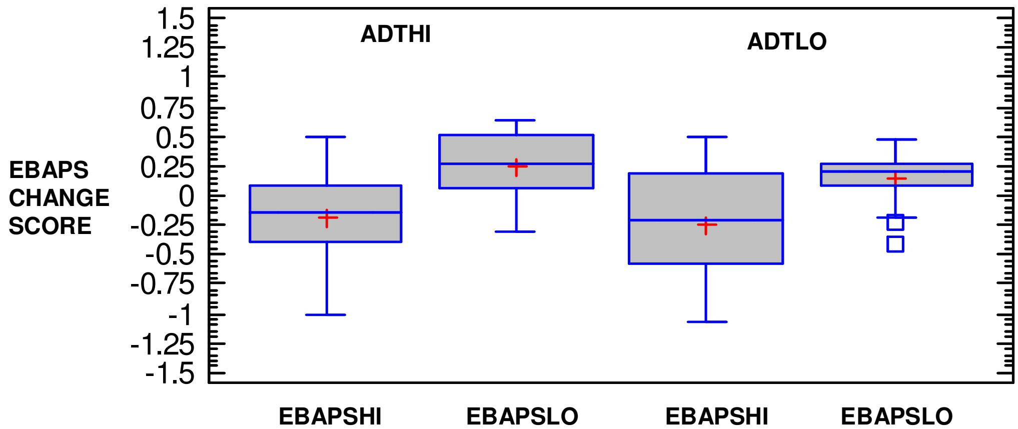


Figure 27 EBAPS change scores for the four groups in SPRING 2007

When looking at the trends in pre- to post-instruction EBAPS changes for WINTER 2007 and SPRING 2007, we see that overall there is no apparent difference between the preEBAPSHI and preEBAPSLO groups for either ADTHI or ADTLO, with one exception.

The median pre- to post-instruction EBAPS change for the ADTHI/preEBAPSLO group in SPRING 2007 appears to be higher than for the similar group in WINTER 2007, and it appears to be the highest median change of all of the groups in SPRING 2007 (see Figure 27). Table 9 shows the EBAPS change scores for our groups in WINTER 2007 and SPRING 2007. We see that the ADTHI/preEBAPSLO group has the highest change score of the eight presented. The change scores for FALL 2006 are omitted since we have discussed the reason that those change scores showed large gains in the preEBAPSLO category.

Table 9 EBAPS change scores for WINTER 2007 and SPRING 2007 for the groups ADTHI/preEBAPSLO, ADTLO/preEBAPSHI, and ADTLO/preEBASPLO. The SPRING 2007 ADTHI/preEBAPSLO change score is the most positive of theses scores.

	EBAPS change			
TERM	ADTHI preEBAPSHI	ADTHI preEBAPSLO	ADTLO preEBAPSHI	ADTLO preEBAPSLO
WINTER 2007	-0.066	-0.20	-0.20	0.11
SPRING 2007	-0.12	0.26*	-0.15	0.11

* $p < 0.01$

This indicates that the instruction intended to foster a change in student epistemologies does have the anticipated effect, since we see that there is a gain for those students with low pre-instruction EBAPS scores in SPRING 2007. In our discussion about content knowledge as quantified by exam scores, it appears that there is a correlation between instruction intended to foster epistemological changes and the exam averages for those students that come in with relatively unsophisticated epistemologies and relatively high general content knowledge - the ADTHI/preEBAPSLO group.

A possible follow-up to our study would identify the ADTHI/preEBAPSLO students and follow them through a term of this instruction in order to see where those changes are coming and what form they are taking. This will be discussed further in the next section.

Discussion

To interpret the findings we described in the preceding section we need to look in detail at the instructional changes whose results they document. We are seeking to understand the implications of our data in evaluating the instructional methods we have introduced. In this section I will discuss the findings that we have from the previous section. This will be a discussion of the EBAPS changes and the content changes.

In prior terms, including FALL 2006 and WINTER 2007 in this study, the labs were nearly independent of the lecture. The instructor (those before me and me as well) would come into the lecture hall and talk about the content for that day. This content may or may not have been relevant to the lab that week. Even in WINTER 2007, when I was using varied instruction techniques, I did not make a complete connection to the labs. This connection happened with the introduction of pre-labs in SPRING 2007.

The previous manuscript showed that student epistemologies were changed through the introduction of the inquiry labs in FALL 2006 compared to Spring 2004. More detailed analysis showed that this is due mostly to the reduction of students with negative EBAPS change scores. The reduction of negative change scores came in two stages; the first stage from Spring 2004 to FALL 2006 and the second stage from WINTER 2007 to SPRING 2007. This is an indication that the connected lecture/lab model was more effective in changing student epistemologies. The important finding is that the inquiry-based labs had the most influence on those students entering with very low EBAPS scores as seen in FALL 2006 change scores.

The exam averages for the 2006-2007 year were not different overall from term to term. The ADT indicated that students in SPRING 2007 entered with significantly lower ADT median scores compared to FALL 2006 and SPRING 2007. The connected lecture/lab instruction managed to help the SPRING 2007 students to gain content knowledge better than the previous two terms. As with the epistemological changes the disadvantaged students were helped by this instruction.

When we looked at the four different groups, we found certain trends that lead to the next important discussion. First, those students best prepare for the course, ADTHI/preEBAPSHI, did not suffer from this instruction. Their exam average was the highest for all three terms.

The second trend that we noticed is that those students not prepared for this course, ADTLO/preEBAPSLO, consistently had the lowest exam average each term. However, again there was a slight but not significant increase in the exam average from FALL 2006 to WINTER 2007 to SPRING 2007. Since the students in SPRING 2007 had significantly lower ADT scores compared to FALL 2006 and WINTER 2007, the connected lecture/lab instructional was successful at bringing the SPRING 2007 ADTLO/preEBAPSLO group to the same content knowledge level as similar groups in the other two terms. Yet, the ADTLO/preEBAPSLO exam average in SPRING 2007 was well below the other exam averages in SPRING 2007.

There is a minimum level of preparation that students need to be successful in this course. If they are in the bottom third of both general astronomy knowledge and epistemology, they will have trouble in the course. What changes in instruction can be made to help these students? It may be that changing the setting is what is needed.

These students may need to be taken out of the lecture/lab model and put into something closely resembling *Workshop Physics*.

We also saw a similar effect for those students in the upper third of both prior general astronomy content knowledge and epistemological beliefs. This is another case of a boundary on the effectiveness of the instruction.

It seems that as far as initial general astronomy knowledge is concerned, the ADTHI/preEBAPSLO groups was prepared for the content, since their median exam average in FALL 2006 and WINTER 2007 is above that of both the ADTLO/preEBAPSHI and ADTLO/preEBAPSLO groups. However, their epistemology of learning science was not sophisticated enough for them to retain the content as well as the ADTHI/preEBAPSHI group did. It was not until the course work made the deeper connection between what I was discussing in the large lecture hall, and what they were doing in the lab, that the ADTHI/preEBAPSLO group was influenced and their median exam average reflects this.

The ADTHI/preEBAPSLO group needed to see the connection of the lab to the lecture in order to be successful. It is unclear how this connection affected their thinking, but it is apparent that it was this connection that made the difference.

The final section of this manuscript will discuss research that can help to answer some of these questions.

Conclusion

In this manuscript we showed that student epistemologies can be positively changed through inquiry-based astronomy instruction. Epistemology-centered instruction does not need to be done at the sacrifice of content gain in the course. We saw that the content gain by students was consistent through the year.

We also saw that a connected lecture/lab model can help those students with less initial general astronomy content knowledge. The exam averages from SPRING 2007 show this. However, for those students with very naïve initial epistemologies, all we needed to positively change their epistemologies was to introduce inquiry-based lab exercises.

In order to extend this discussion, there are three major projects that would be the next step. First, we need to develop an instrument to measure students' epistemologies of science. We saw from the discussion of the EBAPS that the overall Chronbach alpha was 0.79. The EBAPS is measuring a construct overall. However, the subscales that Elby (2001) describes have alphas low enough that we can infer that they are not measuring the same construct. In fact, one of the more important subscales – Real-Life Applicability – has an alpha of 0.04. This indicates that none of the questions are probing the same construct.

The EBAPS measures students' epistemologies of learning science more than their epistemologies of science. From the introductory chapter to this document, we see that it may be difficult to probe students' epistemologies of science, but we may be able to carefully construct and instrument that does this.

A second avenue of research is to identify and follow the ADTHI/preEBAPSLO students. We would like to see how these students are using and internalizing the information they are learning in the course. What makes this group most interesting is that they have high general content knowledge, yet they still have a naïve epistemology when entering the course. Interviews of these students would help to understand how they think about science.

The third avenue of research is a deeper look into the demographics. The demographic data that I examined using the ADT was informative in allowing us to compare the students in all three terms. However, those demographics were too general. In order to identify those students most affected by my instructional model, I would like the categories to be more finely segregated. In preparing the ADT, Hufnagel (2002) and the other developers of chose to consolidate Humanities, Social Sciences, and Art into one major group, but I would argue that these are three very different groups. Another category where finer grained separation would be advised is age. It is highly unlikely that students from age 0-16 are attending college. My choice to use the demographics from the ADT was predicated by my use of the ADT. I felt it unnecessary at the time to administer another demographic test after I have collected that information once.

CHAPTER FOUR

Discussion

In the introductory chapter and the previous two manuscripts, I made the case for a redesigned non-major introductory astronomy course. The main factors that I discussed were:

- Each year there are upwards of 250,000 students enrolled in a course like the one described here.
- This is often the only science course many students will take in college.
- There is an identified need for non-major courses to help students understand the process and nature of knowledge in science.
- Non-major introductory science courses can be used as a tool to recruit undecided freshmen into a science major.
- Courses that do not focus an epistemological instruction, while increasing content knowledge, decrease student epistemologies of learning science.
- This is an area that has not been studied at this level in astronomy education.

A synopsis of the previous manuscripts follows. First I will review the epistemological and instructional frameworks that I used. I will follow this with a discussion of the findings from the data collected in this study connecting to the framework.

In the development of my framework, the first main point that I raised was the difference in epistemology of science and epistemology of learning science. The epistemology of science – science is tentative, science is subjective, science is cultural, *etc*. – may be too sophisticated and complex for novice science students. Students may hold a set of beliefs about the epistemology of science, but they can only articulate those beliefs in the context of their own experiences. Thus students may voice their epistemologies of science through their epistemologies of learning science. The instruments discussed in both manuscripts one and two –EBAPS and MPEX – are more closely aligned with measuring students' epistemologies of learning than with measuring students' epistemologies of science. My framework discussion is based on this assumption. My data analysis and connection to instruction also takes this into account.

In the introductory chapter, I discussed the epistemological framework that I am using. It is based on the work of Schommer (1989), who found that student epistemologies are a set of coherent independently developed beliefs. A student may hold a naïve belief structure and a sophisticated belief structure simultaneously. As an example, consider the student that thinks of knowledge as simple. This student believes that knowledge is made of individual unconnected facts. This is a naïve epistemology. However, this student may also think that learning in science does not depend on innate ability, but anyone with enough work can learn science. This is a sophisticated epistemology. Both exist in the student, because each set of beliefs is independent.

Instead of using Schommer's complete framework, I connected her ideas to those of diSessa (1993), Hammer and Elby (2001), and Louca *et al.* (2004). In the epistemological models of diSessa (1993), Hammer and Elby (2001), and Louca *et al.* (2004), student epistemologies are fine-grained cognitive resources that are contextually activated. They argue that students will not hold coherent belief systems, but all beliefs are possible, and whether a student will be naïve or sophisticated depends on which resource is activated. In my framework, I have suggested that Schommer (1989) is correct in framing epistemologies as a set of beliefs, but I contend that these beliefs are contextually dependent. Therefore a student can have sophisticated epistemologies about how physics knowledge is constructed, but have naïve beliefs about how biology knowledge is constructed. To the professional scientist, these differences in physics and biology are the contexts of the science, but the methods for acquiring and using knowledge are the same. To students in college these may seem like completely different topics, and they need to be approached differently. A study of my framework, though interesting, is outside the scope of this study.

The third item that I introduced in the opening chapter is a discussion of scaffolded instruction as my main framework for the instructional design. Peer Instruction (Mazur, 1997) was the concrete example of scaffolded instruction in the large lecture classroom, and in the first manuscript the inquiry cycle of White and Frederiksen (1998) is the example of scaffolded instruction in the lab class. The main goal of my instruction model is to effect positive changes in students' epistemology of science. In order to accomplish this I need to help students change the beliefs about science that they initially bring into the classroom. My chosen means is through scaffolded instruction.

My model of instruction was implemented over the course of three terms in the 2006-2007 academic year. The first piece introduced was lab exercises based on the inquiry cycle of White and Frederiksen (1998). Initially the labs encompassed the entire cycle. This led to problems with students completing the labs in the designated two hours each week. There were gains in post-instruction EBAPS scores compared to the baseline measurements of Spring 2004. In WINTER 2007, I introduced varied instructional methods in the lecture portion of the course. Instead of relying on only PI, I also used concept maps similar to Zeilik *et al.* (1997) and tutorials similar to Prather *et al.* (2004). In WINTER 2007 the format of lab exercises did not change. We still ran into the problem of students not completing the exercises in the labs, but the post-instruction EBAPS results showed another gain over FALL 2006. The final stage of the complete instructional model took the labs and broke the inquiry cycle into two parts; a part in the lecture hall, introduced as pre-labs, and a part in the lab class. This created connectivity between the lecture and lab that had been absent in prior terms. The final instruction model was introduced in SPRING 2007.

The final model of instruction showed positive indicators of epistemological change. When looking at the change score on the EBAPS, post-instruction minus pre-instruction, I noticed that the number of students with negative change decreasing from term to term. This indicated that the full model of instruction made positive changes in student epistemologies.

One item discovered in the data analysis was noted in the second manuscript: that FALL 2006 had a large population of students with very low pre-instruction EBAPS scores, classified as very naïve. This group of students showed the highest change score in all of the terms. These results indicated that the inquiry cycle of the labs might be a factor in changing student epistemologies. The post-instruction

EBAPS of the FALL 2006 group indicated that this had the effect of helping the naïve students think more like their sophisticated counterparts.

While building an epistemologically based instructional model, I needed to make sure that my students were also learning the content that I was teaching. The second analysis of manuscript two was dedicated to this analysis.

The first aspect of the instruction that I noted was that epistemologically instruction as it was carried out in SPRING 2007 resulted in exam averages that were equivalent to those students in FALL 2006 and WINTER 2007 (Kruskal-Wallis, p = 0.33). However, the students entering SPRING 2007 had significantly lower scores on the ADT than those students of the previous two terms (ANOVA with Scheffé intervals, $p < 0.05$). My analysis indicated that students without good general astronomy content knowledge could learn and retain the content at a level similar to those with a higher median ADT score. This shows that my instructional method could change epistemology and improve content gain in the same instruction.

In manuscript two I also presented findings on four major groups that were defined by ADT and pre-instruction EBAPS. Those groups were ADTHI/preEBAPSHI, ADTHI/preEBAPSLO, ADTLO/preEBAPSHI, and ADTLO/preEBAPSLO. Results indicated that students in the ADTHI/preEBAPSHI group consistently had higher median exam averages. We also noted that the ADTLO/preEBAPSLO group consistently had the lowest exam averages. The other two groups had median exam averages between the ADTHI/preEBAPSHI and ADTLO/preEBAPSLO groups with two exceptions – the ADTLO/preEBAPSHI group in FALL 2006 and the ADTHI/preEBAPSLO group in SPRING 2007.

The group that showed significant change in their median exam scores from term to term was the ADTHI/preEBAPSLO group. This group did not have a median exam score significantly higher than the ADTLO/preEBAPSHI group (p <0.05) in FALL 2006 or WINTER 2007. However, the ADTHI/preEBAPSLO group had median exam scores the same as the ADTHI/preEBAPSHI group in SPRING 2007. The higher median exam scores for the ADTHI/preEBAPSLO group correlated to higher median EBAPS change score from term to term (Kruskal-Wallis, $p < 0.05$). Though the EBAPS change score for the ADTHI/preEBAPSLO group in SPRING 2007 was not significantly greater than the EBAPS change score for the other three groups in SPRING 2007 (Kruskal-Wallis, $p < 0.05$), the median is the highest among the four groups.

The previous analysis of the two manuscripts shows that our fully integrated instruction method in SPRING 2007 had an effect on the epistemology and content gain of the students in SPRING 2007. They entered with much less content knowledge and had median exam scores similar to FALL 2006 and WINTER 2007. Elby (2001) showed that instruction intended to foster epistemological gains was successful at giving students a deeper understanding of how they learn science. Redish *et al.* (1998) also showed that instruction and curricula designed to foster content knowledge created a regression in epistemology after instruction. In my project, I showed that one can maintain content gain while fostering epistemological progress in students. However, what epistemology have we changed? Have we only changed the students' epistemologies of learning science or have we also changed their epistemologies of science? The next section makes connections between both in order to discuss the need for further instrument development in this field.

Attempting to evaluate novice science students on a construct as complicated as the epistemology of science may be significantly more difficult than measuring their epistemologies of learning science. Where some science experts, building on many years of experience, have a sound sophisticated system of beliefs on what it means to build knowledge in science, novice students may have only their experiences in the science classroom to draw upon. If this case, how does one get to an understanding of how students view the nature of science? It may be that for us to understand what epistemology of science a student has, we must actually measure something different. In asking a student how she thinks she personally constructs knowledge in science, we might also get some insight into what she believes about how knowledge is constructed in science. In short, to understand a novice's epistemology about science, we need to probe a novice's epistemology about how they learn science.

This is more in line with what the Epistemological Beliefs Assessment for Physical Sciences (EBAPS) actually does. Let us look at the following example in detail. Suppose a student answers "agree" the following question from EBAPS:

> 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 (Elby, 2001).

This would be reported as a naïve epistemology about learning physics. But if we look closely at this question, we see that there might be other deeper meaning that we can pull out of this answer. It appears that the student sees physics problem solving as connecting the given problem to the right equation. Hammer (1994) showed us this in his study of physics students. The concepts underlying the problem seem to interfere with the problem solving process. The goal of problem solving is to arrive at an answer. The epistemology of science that we can infer from this student is that the student may not understand the tentative aspect of the epistemology of science.

There is a jump from agreeing to the EBAPS question to the tentative nature of science, but this can be seen in the work of Hammer (1994). In his study, David Hammer (1994) showed that the student Liza worked on problems to arrive at the answer. The answer was either correct or incorrect. We can assume that if she did not get the right answer she would assume that she had done something incorrect and try again. There was no reflection about why she misunderstood the problem. She wanted to know which equation would give her the right answer. This would lead to the idea that the right answer was important and the right equation for that problem was important. This gets us to the tentative nature of science. Liza would assume that physicists are always looking for the right answer. If they are arguing about the answer to a “problem”, then they have not applied the right equation to it. These students have been shown that discovering nature is about finding the answer to what nature is showing us. In Hammer’s study this is not entirely the fault of Liza, since we saw Ellen abandoned the more sophisticated point –of view in order to become more successful in the course. Hammer (1994) implies that this may be more typical of college physics courses than we as instructors are willing to admit.

This entire previous discussion attempted to make the connection from the questions on the MPEX or EBAPS to the epistemology of science. The questions on these two instruments seem to ask questions about students' epistemologies of learning science. The epistemology of science may be too complicated to ask of novice students. However, there may be a way for us to ask students questions about their epistemologies of science. The EBAPS gives us a roadmap. The questions from the EBAPS include not only straight phrases to agree or disagree with, but stories about interactions between two people for which students must choose a side to the argument. This may be the necessary format to probe student epistemologies of science. I will discuss this more in the further research section next.

Further Research

In the previous sections of this discussion, I mentioned that there is a need to develop a better understanding of student epistemologies of science. There are other areas of research that are also important for the continuation of this research. I will discuss the important areas that I see as reasonable next phases of this research.

First, we would like to identify groups of students that enter the non-major introductory astronomy with naïve epistemologies. In the discussion, we identified a group with naïve epistemologies, but we had little information on their demographics. The demographic data that I examined using the ADT was informative in allowing us to compare the students in all three terms. However, those demographics were too general. In order to identify those students most affected by my instructional model, I would like the categories to be more finely segregated. In preparing the ADT, Hufnagel (2002) and the other developers chose to consolidate Humanities, Social Sciences, and Art into one major group, but I would argue that these are three very different groups. Another category where finer grained separation would be useful is age. It is highly unlikely that students from age 0-16 are attending college. My choice to use the demographics from the ADT was dictated by my use of the ADT. I felt it unnecessary at the time to administer another demographic test after I had collected much of the information once.

Second, in order to understand students' epistemologies of science, we need a different instrument. We need to create a simple discussion of the epistemology of science in order to probe the students' beliefs. This would be something similar to the EBAPS in form, but the questions would be geared toward epistemology of science. In designing the EBAPS, Elby (2001) created questions using arguments between two people. Students commit to a position by answering who they most agree with. Their position determines their view on the question.

If my framework of epistemology is useful, student epistemologies are context sensitive. The lecture and the labs constitute different contexts for learning astronomy. It would be useful to develop methods that could measure how student epistemologies are being affected in the lecture versus how they are being affected in the lab. This would entail the use of audio, video, and interviews. These were not part of the previous study, but should be included in the next stage of research.

In the previous document we discussed changing student epistemology through scaffolded instruction. Scaffolded instruction in the connected lecture/lab improved student epistemologies as quantified by the EBAPS. As Zeilik *et al.* (1998) claims, the students that take the non-major introductory astronomy course may have this as their only science course. These students will make decisions based on their understanding of science. As instructors we need to make sure our students leave with a better understanding of the nature of science. In order to do this we must develop their epistemology of science.

References

Abd-El-Khalick, F. (2004). Over and over again: College students' views on the nature of science. In Flick, L.B., and Lederman, N.G., (eds.) *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Adams, J. P., Prather, E. E., & Slater, T. F. 2005, *Lecture-Tutorials for Introductory Astronomy*, First Ed., Upper Saddle River, NJ: Prentice Hall.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., and Schauble, L. (2003). Design experiments in educational research. *Educational Researcher,* 32(1), 9-13.

Crouch, H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9), 970-977.

diSessa, A.A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 10, 105-225.

Edelson, D.C. (2002). Design research: What do we learn when we engage in design. *The Journal of the Learning Sciences,* 11(1), 105-121.

Elby, A. (2001). Helping physics students learn how to learn. *Physics Education Research, American Journal of Physics Supplement,* 69(7), S54-S64.

Elby, A. (2001). *The idea behind EBAPS*. Retrieved November, 2004, http://www2.physics.umd.edu/~elby/EBAPS/idea.htm.

Fraknoi, A. (2001). Enrollments in Astronomy 101 Courses. *Astronomy Education Review,* 1(1), 121-123.

Gobert, J., and Discenna, J. (1997). *The relationship between students' epistemologies and model-based reasoning*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Green, P.J., (2002). *Peer instruction for astronomy.* Upper Saddle River, NJ: Benjamin Cummings.

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(1), 64-74

Hammer, D. (1989). Two approaches to learning physics. *The Physics Teacher,* 27, 664-670.

Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction,* 12(2), 151-183.

Hammer, D., and Elby, A. (2002). On the form of personal epistemology. In Hofer, B.K., and Pintrich , P.R. (eds.) *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mawah, NJ: Lawrence Erlbaum Associates, Publishers.

Hammer, D., and Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of Learning Sciences,* 12(3), 53-90.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30, 141-158.

Hofer, B.K. (2002). Personal Epistemology as a psychological and educational construct: An introduction. In Hofer, B.K., and Pintrich , P.R. (eds.) *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mawah, NJ: Lawrence Erlbaum Associates, Publishers.

Hogan, K. & Pressley, M. (1997). *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline Books, Inc

Hufnagel, B. (2002). Development of the Astronomy Diagnostic Test*, Astronomy Education Review*, 1(1) 47-51.

Laws, P.W. (1996). Millikan lecture 1996: Promoting active learning based on physics education research in introductory physics courses. *American Journal of Physics,* 65(1), 14-21.

Lederman, N.G. (2004). Syntax of nature of science within inquiry and science instruction. In Flick, L.B., and Lederman, N.G., (eds.) *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lent, R.W., Lopez, F.G., and Bieschke, K.J. (1991). Mathematics self-efficacy: sources and relation to science-based career choice. *Journal of Counseling Psychology,* 38(4), 424-430.

Lising, L., and Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *The American Journal of Physics,* 73(4), 372-382.

Louca, L., Elby, A., and Hammer, D., and Kagey, T. (2004). Epistemological resources: Applying a new framework to science instruction. *Educational Psychologist,* 39(1), 57-68.

Luzzo, D.A., Hasper, P., Albert, K.A., Bibby, M.A., and Martinelli, E.A. Jr. (1999). Effects of self-efficacy interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of Counseling Psychology,* 46(2), 233-243.

Manogue, C., Siemens, P., Tate, J., Browne, K., Niess, M.L., & Wolfer, A.J.,(2001) Paradigms in Physics: A New Upper-Division Curriculum, *American Journal of Physics* **69**, 978-990

Marzano, R.J., Pickering, D.J., and Pollock, J.E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement.* Upper Saddle River, NJ: Pearson, Merrill Prentice Hall.

Mayer, R. E. (2002). *The promise of educational psychology, Volume II: Teaching for meaningful learning*. Upper Saddle River, NJ: Pearson Education, Inc.

Mazur, E. (1997). *Peer Instruction; A user's Manual.* Upper Saddle River, NJ: Prentice-Hall, Inc.

McDermott, L.C., and Redish, E.F. (1999). Resource Letter: PER-1: Physics education research. *American Journal of Physics*, 67, 755-767.

Merriam-Webster Online Dictionary. 2008

Meyer, D.K., and Turner, J.C. (2002). Using Instructional Discourse Analysis to Study the Scaffolding of Student Self-Regulation. *Educational Psychologist,* 37(1), 17-25.

Moore, W.S. (2002) Understanding learning in a postmodern world: Reconsidering the Perry scheme of ethical and intellectual development. In Hofer, B.K., And Pintrich , P.R. (eds.) *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mawah, NJ: Lawrence Erlbaum Associates, Publishers.

Mulvey, Patrick & Nicholson, Starr 2001, Enrollments and Degrees Report, AIP Report, Number R-151.37

National Science Foundation. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology* (NSF 96-139). Washington, DC: The Author.

National Research Council (1999). *Transforming undergraduate education in science, mathematics, engineering, and technology.* Washington, DC: National Academy Press.

Palinscar, A.S. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist,* 21, 73-98.

Prather, E.E., Slater, T.F., Adams, J.P., Bailey, J.M., Jones, L.V., Dostal, J.A. (2004). Research on a Lecture-Tutorial Approach to Teaching Introductory Astronomy for Non–Science Majors. *Astronomy Education Review,* 2(3), 122-136.

Prather, E.E., Slater, T.F., Adams, J.P., and Brissenden, G. (2008). *Lecture tutorials for introductory astronomy.* San Francisco, CA. Pearson Addison-Wesley.

Rebello, N.S., and Zollman, D.A. (2005). Trends in physics education research – a personal perspective. Unpublished.

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American Journal of Physics*, 66, 212-224.

Ryder, J., Leach, J., and Driver, R. (1999). Undergraduate Science Students' Images of Science. *Journal of Research in Science Teaching*, 36(2), 201–219.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498-504.

Schommer-Aikens, M. (2002). An evolving theoretical framework for an epistemological belief system. In Hofer, B.K., and Pintrich , P.R. (eds.) *Personal epistemology: The psychology of beliefs about knowledge and knowing.* Mawah, NJ: Lawrence Erlbaum Associates, Publishers.

Wiggins, G., and Mctighe, J. (1998). *Understanding by Design*. Upper Saddle River, New Jersey: Merrill-Prentice Hall.

White, B. Y., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction.* 16 (1), 3-118.

Wood, D., Bruner, J., and Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17 , 89–100.

Zeilik, M., Schau, C., Mattern, N., Hall, S., and Teague, K.W. (1997). Conceptual astronomy: A novel model for teaching postsecondary science courses. *American Journal of Physics,* 65, 987-996.

Zeilik, M., Schau, C., & Mattern, N. (1999). Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters. *American Journal of Physics,* 67(10), 923-927.