A COMPARATIVE STUDY OF THE COGNITIVE AND METACOGNITIVE DIFFERENCES BETWEEN MODELING AND NON-MODELING HIGH SCHOOL PHYSICS STUDENTS

By

Kathy L. Malone

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology
Center for Innovation in Learning
Carnegie Mellon University
Pittsburgh, PA

2006

Thesis Committee
Dr. Ruth Chabay, Chairperson of Committee
Dr. John Hayes
Dr. Marsha Lovett
Dr. Bruce Sherwood

Date:

May 1, 2006

ABSTRACT

A COMPARATIVE STUDY OF THE COGNITIVE AND METACOGNITIVE DIFFERENCES BETWEEN MODELING AND NON-MODELING HIGH SCHOOL PHYSICS STUDENTS

The Modeling Instruction pedagogy for the teaching of physics has been proven to be quite effective at increasing the conceptual understanding and problem-solving abilities of students, as measured by the Force Concept Inventory and the Mechanics Baseline Test, to a much greater extent than that of non-modeling students. Little research has been conducted concerning the cognitive and metacognitive skills that modeling students develop that allow for these increases. In this thesis, two studies were designed to answer the following question:

> In what ways do the knowledge structures, metacognitive skills, and problem-solving abilities differ between modeling and non-modeling students?

In Study 1, the knowledge structures developed by both modeling and non-modeling students were determined using a card sort task. The type of knowledge structure developed by the students was quantified via expert, surface feature and "questions asked" scores. The student's knowledge structures were then correlated to the scores they obtained on two measures: FCI and a problem solving task (PS Task). It was discovered that at the end of a year long course that the modeling students had a more expert-like knowledge structure while non-modeling students produced structures that were less expert-like and more surface feature or "questions asked" oriented. In addition, the expert score correlated highly with performance on both the FCI and PS Task scores demonstrating that a higher expert score predicted a higher value on each of these measures while a higher surface feature score predicted a lower score on both of these measures.

In Study 2, a verbal protocol design allowed for a detailed study of the problem-solving and metacognitive skills utilized by the two groups. It was determined that the skills developed in both of these areas by the modeling students were more expert-like as based upon prior research. In addition, the modeling students produced significantly fewer physics errors while catching and repairing a greater percentage of their errors. This study also allowed for an exploration of the shift in metacognitive behaviors before, during and after an error correction episode.

Keywords: Modeling Method, physics, knowledge structures, problem-solving, metacognition, self explanation, Force Concept Inventory, errors

This thesis is dedicated to my husband, Dan, and my children, Victoria, Anna and Daniel Lee.

ACKNOWLEDGEMENTS

I would like to thank all of my committee members for their guidance. My advisor, Ruth Chabay, has helped me in innumerable ways over the past several years. She always made sure that my analysis stayed focused on the physics abilities of the students. I am also grateful for Bruce Sherwood's help in problem design and consistently holding me accountable for realistic problem sets. In addition, Ruth and Bruce constantly opened their home and pantry in order that I might visit them in North Carolina to review my work. Marsha Lovett's guidance during the development of my coding scheme for Study 2 was indispensable. Finally, I would like to thank Dick Hayes. Without Dick's day to day guidance, encouragement and general hand holding I might not have been able to finish this thesis.

Most importantly I would like to thank by husband, Dan Malone, and my children. Dan's support, encouragement and proof reading help not to mention his willingness to take over innumerable day to day chores while making sure that I had at least one quite day at home every weekend in which to work have been greatly appreciated. Without his support this thesis may not have ever been finished.

TABLE OF CONTENTS

Chapter 1
Introduction

Prior to the late 1950's high school physics was considered by most students a course scheduled only as a last resort. The idea of having students think like physicists was not considered an option. This state of affairs lasted until 1957 when the Soviet Union launched Sputnik I. The United States government spent a considerable amount of money to not only put Americans into space but also to improve science education especially physics education. The main thrust of the efforts was to produce a greater number of students able to become contributing scientists. The new government funding led to the production of new curriculum movements which stressed hands-on student involvement in all areas of science education. In the area of physics, the Physical Science Study Committee developed a hands-on physics curriculum simply called PSSC Physics and a group from Harvard developed Project Physics which also contained a historical component. These curriculum projects emphasized hands-on educational approaches and were meant to allow physics to be studied by more than just the countries' most promising students as was the case prior to Sputnik. The emphasis on hands-on educational techniques lasted well into the seventies with people assuming that our students were learning physics to a greater extent than ever and should be able to continue towards promising careers in science. In the 1980's physicists conducting educational research discovered, to the disbelief of many, that these students had significant gaps in their understanding of physics. It seemed that our new courses in high school and college might not be producing the exceptional physics students we had hoped for. It seemed that the money spent to generate new physics high school physics curriculum's such as PSSC was for naught. The research conducted in the 1980's demonstrated that students had a number of misconceptions about physics and that these misconceptions continued to be in place even after traditional instruction (Andersson and Karrqvist, 1983; Caramazza, McCLoskey and Green, 1981; Champagne, Klopfer and Anderson, 1980; Clement, 1982; Cohen, Eylon, and Ganiel, 1983; DiSessa, 1982; Goldberg and Anderson, 1989; Goldberg and McDermott, 1986, 1987; Gunstone, 1987; Gunstone and White, 1981; Halloun and Hestenes, 1985a, 1985b; Lawson and McDermott, 1987; Maloney, 1984; McCloskey, 1983; McCloskey, Caramazza, and Green, 1980; Trowbridge and McDermott, 1980, 1981; Viennot, 1979; White, 1983). This research allowed us to understand what pieces of knowledge the students were missing after a course or two in physics but did provide any information about deficiencies in the exiting student's knowledge organization and their problem-solving skills.

Cognitive psychologists during the same time period were conducting research that enhanced our understanding of the knowledge organization of novice physics students and expert physicists (Chi, Feltovich and Glaser, 1981; Hardiman, Dufresne, and Mestre, 1989; and Reif and Allen, 1992). This research demonstrated that novice students' knowledge was fragmented and poorly organized (Chi et al., 1981 and Chi, Glaser, and Rees, 1982). Research initiatives were also designed that analyzed the problem-solving capabilities of novice students in relation to the skills of experts in the field (Simon and Simon, 1978; Simon and Simon, 1979; Larkin, 1979; and Larkin, McDermott, Simon and Simon, 1980a, 1980b). The novice subjects in these studies had typically completed a year of traditional college physics instruction. The cognitive state and problem-solving capabilities of the traditional high school age physics student is probably less developed than the students used in this study.

The performance of novices in all of the studies above caused many to pause and consider how these physics students could be so deficient in knowledge, organization and problem-solving skills after a year-long effort at instruction. Since the implicit understanding is that instruction should improve the understanding and knowledge level of our students by guiding them towards greater expertise, this was a problematic state of affairs. This research inspired the development of a number of reformed courses in physics such as Minds-On Physics, Physics Resources and Instructional Strategies for Motivating Students (PRISMS), Active Physics, Constructing Physics Understanding (CPU) and Modeling Instruction. The research mentioned above gave the researchers an understanding of the conceptions the students were missing and resulted in the production of numerous activities that facilitated the incorporation of these missing conceptions. A number of the new reformed courses such as Modeling Instruction, CPU and Minds-on Physics also were influenced by the work of cognitive psychologists on knowledge structures.

This thesis deals specifically with the Modeling Instruction Pedagogy; therefore, one must ask: What are the basic tenets of Modeling Instruction? The active processing of information by students vs. passively receiving information has been shown necessary for learning to occur (Brown, 1994; Chi, Bassok, Lewis, Reimann, and Glaser, 1989; Kintsch, 1993). Constructivist reform efforts are based upon the idea that students actively construct their own knowledge (Piaget, 1963, 1970; Vygotsky, 1962). The need for active engagement is not farfetched and has been proven to be quite successful in a number of fields (Collins, Brown, and Newman, 1989; Hake, 1998; Kahn and Carver, 1988, Reif, Larkin and Brackett, 1976). Indeed, the modeling method utilized in high school was built on the premise that active engagement is necessary and as such it is constructivist based (Wells, Hestenes, and Swackhamer, 1995). In addition, Modeling is based upon a modified learning cycle similar to the learning cycle advocated by Robert Karplus (1977). A synopsis of the modeling method, its objectives and instructional strategies is shown in Figure 1-1. The modeling classes described in this thesis are taught based upon the synopsis in Figure 1-1.

The Modeling Method aims to correct many weaknesses of the traditional lecture-demonstration method, including the fragmentation of knowledge, student passivity, and the persistence of naïve beliefs about the physical world.

Coherent instructional objectives

- To engage students in understanding the physical world by ***constructing and using scientific models*** to describe, to explain, to predict, to design and control physical phenomena.
- To provide students with ***basic conceptual tools*** for modeling physical objects and processes, especially mathematical, graphical and diagrammatic representations.
- To familiarize students with a small set of basic models as the ***content core*** of physics.
- To develop insight into the ***structure*** of scientific knowledge by examining how *models* fit into *theories.*
- To show how scientific knowledge is ***validated*** by engaging students in *evaluating* scientific models through comparison with empirical data.
- To develop skill in all aspects of modeling as the ***procedural core*** of scientific knowledge.

Student-centered instructional design

- Instruction is organized into ***modeling cycles*** which engage students in all phases of model development, evaluation and application in concrete situations — thus promoting an integrated understanding of modeling processes and acquisition of coordinated modeling skills.
- The teacher sets the stage for student activities, typically with a demonstration and class discussion to establish common understanding of a question to be asked of nature. Then, in small groups, students ***collaborate*** in planning and conducting experiments to answer or clarify the question.

- Students are required to present and justify their conclusions in oral and/or written form, including a ***formulation*** of models for the phenomena in question and ***evaluation*** of the models by comparison with data.
- Technical terms and representational tools are introduced by the teacher as they are needed to sharpen models, facilitate modeling activities and improve the quality of discourse.
- The teacher is prepared with a definite ***agenda*** for student progress and ***guides*** student inquiry and discussion in that direction with "Socratic" questioning and remarks.
- The teacher is equipped with **taxonomy** of typical student misconceptions to be addressed as students are induced to articulate, analyze and justify their personal beliefs.

A modeling approach to instruction was adopted because models are fundamental to science and its practice (Wells et al. 1995). The modeling cycle involves model development via a paradigm lab and the deployment of the model produced in a number of different circumstances. During the paradigm lab and its ensuing post-lab, the model which describes a physical system is articulated and validated. The final model developed consists of several representations: verbal, diagrammatic, algebraic or mathematical and graphical representation. This model is then deployed by the students. The model deployment consists of utilizing the model developed in new situations in order to free it from the context of the paradigm lab. The new deployment situations can take the form of additional lab experiments and more traditional problem sets. The modeling mechanics course allows for the students to develop a basic set of models shown in Figure 1-2.

Kinematical Models	Causal Models
Constant Velocity	Free Particle
Constant Acceleration	Constant Force
Uniform Circular Motion	Central Force
Collision	Impulsive Force

Figure 1-2: Basic Models developed in the Mechanics Modeling Course (adapted from Hestenes, 1996)

During initial development of the model each lab group presents its initial model of the physical system it is experimenting with and its associated representations to the rest of the class. The lab group must have evidence from the lab to back up their initial model design. The ensuing discussion is quite intense with the teacher guiding the discussion and upon occasion asking questions of the lab groups using Socratic dialogue. After the lab discussion the teacher helps the class to reach consensus on the basic model representations. After consensus has been reached the class moves into the next part of the cycle, model deployment. During model deployment student groups are given problems with which to practice deploying the model. The groups complete the problems on fairly large sections of whiteboards so that the solution method chosen can be presented to the rest of the class. During the presentation the student group is questioned about their methods by the rest of the class and the teacher. The solution approaches routinely require the use of multiple representations of the model. In this fashion the students constantly view alternative solutions which utilize the different model representations. For example, they may observe the same or similar problem being solved with the use of an algebraic representation in one instance while on the next whiteboard the problem has been solved with a graphical representation. As the cycle continues the students deepen their understanding of the models and became independent thinkers who can easily defend their points of view (Wells et al. 1995). Within the class you commonly hear students being asked: "Why do you think that"; How do you know that?" and "Does that answer make sense?".

The reliance on models should not only allow the students to emulate scientific practice but it should also produce a number of other benefits for the students. The developers of the modeling pedagogy believe that this reliance on the development of basic physics models "should help the students develop a more coherent, flexible and systematic understanding of physics" (Wells et al 1995).

The Modeling Instruction method in physics has been tested extensively utilizing the two tests designed for this purpose. The developers of the Modeling pedagogy developed a diagnostic test, the Force Concept Inventory (FCI), which could be used to test the physics knowledge (i.e., misconceptions) of the students (Hestenes, Wells and Swackhamer, 1992; Halloun and Hestenes, 1985a). A companion test, the Mechanics Baseline test (MBT), was designed to test the problem-solving skills of the exiting student (Hestenes and Wells, 1992). The FCI and MBT only test for single particle systems. The tests cover basic kinematics and dynamics but do not include rotational or energy concepts. Since the inception of these tests, they have been used extensively in physics education not only to test the efficacy of the Modeling Pedagogy but numerous other curriculum designs as well (Hake, 1998; Cummings, Marx, Thornton, and Kuhl, 1999; Cheng, Thacker and Cardenas, 2004; Kalman, Rohar, and Wells, 2004; Morote and Pritchard, 2002; etc.). The efficacy of the Modeling Instruction Pedagogy has been well established as a number of studies have shown that the method helps to produce gains in understanding well beyond that of traditional methods (Hake, 1998; Hestenes, Wells and Swackhamer, 1992; Hestenes and Wells, 1992; Vesenka, Beach, Munoz, Judd and Key, 2002). The positive results of these assessments led the Department of Education to recognize the pedagogy and list it in the National Dissemination Network.

Prior research has shown that modeling students have a stronger knowledge base than traditional students and this base is correlated to better problem-solving performance on their part (Hestenes et al., 1992; Hestenes and Wells, 1992; and Hake, 1998). However, the FCI and the MBT focus on a restricted segment of the course domain and only describe a small portion of the exiting student's abilities, behaviors and conceptions. For example, no research has been completed on the cognitive structures the students possess and the problem-solving behaviors (both cognitive and metacognitive) they develop that allow them to become better problem solvers. In fact, no attempt has been made to determine if the exiting students do actually develop a more coherent knowledge structure after a year-long course via Modeling Instruction or in any of the other reform curriculums. The lack of studies concerning the cognition of modeling students seems to be a gap in the research that should be connecting the two fields of physics education and cognitive psychology. Given this lack it is no wonder that there has been little research that correlates the more coherent "expert-like" knowledge structures to improved problem solving ability. This thesis attempts to obtain a fuller picture of why the modeling student exhibits more enhanced qualitative and quantitative abilities. The question addressed in this thesis is: In what ways do the knowledge structures, metacognitive skills, and problem solving abilities differ between modeling and non-modeling students?

This thesis details my efforts to determine the final conceptual knowledge structure of the modeling student and to evaluate the problem-solving strategies they develop after a year-long course in comparison to that of non-modeling students. Chapter two reviews prior literature in the following areas: documented proof of modeling efficacy, metacognition, knowledge structures and schemas, problem solving strategies and studies that link the three cognitive issues together. Chapter three describes the results of a study designed to determine the knowledge structure of modeling and non-modeling students and correlates these results to their conceptual understanding and problem-solving abilities. Chapter four describes the results of a talk-aloud protocol study designed to determine the differences in the processes that the two experimental groups utilize when solving physics problems and their ability to recover from physics errors. Chapter five presents a summary of all of the results, suggestions for educational design improvements and concludes with suggestions for future research.

Chapter 2

The Convergence of Knowledge Organization, Problem-Solving Behavior, and Metacognition Research with the Modeling Method of Physics Instruction

2.1 Introduction

In this thesis I will be investigating the differences in conceptual understanding, problem solving, metacognitive behaviors and knowledge organization of non-modeling and modeling high school students. This chapter will discuss the previous research pertinent to these areas. The chapter is divided into the following five sections: modeling efficacy research; problem-solving behaviors and strategies; knowledge organization and schemas; metacognitive behaviors and strategies and the self-explanation effect. At times there is considerable overlap in some areas. In addition, research articles in each section are separated by studies that are exploratory in nature vs. studies that researched the efficacy of methods developed to improve student learning based upon the previous exploratory studies.

2.2 The Efficacy of the Modeling Instruction Pedagogy

As mentioned in Chapter 1 the Modeling Pedagogy is one of the few physics reform programs that have been substantially shown to improve students' conceptual understanding of physics and their ability to solve problems. Since the majority of misconception research dealt with isolated concepts, Halloun and Hestenes (1985a) decided to design an instrument that would allow one to assess the knowledge of students before and after physics instruction specifically in the area of the force concept. The subsequent instrument, called the Mechanics Diagnostic Test (MDT), focused on concepts shown to be deficient in the previous misconceptions research in the domain of mechanics. A later version of the MDT was called the Force Concept Inventory (FCI). The MDT was written in language that students without physics training could easily understand. The instrument was administered to college and high school students both pre and post instruction. Halloun and Hestenes (1985a) discovered that the qualitative knowledge gain in conventional physics instruction was extremely poor and independent of the professor. This meant that at the end of the instruction not only were basic Newtonian concepts lacking but misconceptions about mechanics remained firmly in place. In subsequent research Halloun and Hestenes (1985b) were able to develop a taxonomy of common sense beliefs which was based upon item selection on the MDT and student interviews. They classified these alternative beliefs in terms of specific Newtonian concepts so that the taxonomy could act as a guide when assessing instructional interventions.

The MDT clearly demonstrated that there was a need for the development of a radically different teaching approach that would help students to develop a clearer understanding of Newtonian concepts and help to remove their misconceptions. Halloun and Hestenes developed an instructional intervention centered on model-based reasoning that could improve students' grasp of Newtonian concepts. Hestenes (1987, 1992) argued that an analysis of the structure of scientific knowledge leads one to identify that development and deployment of models is the main activity of scientists. The models in mechanics are highly developed and can provide a coherent structure that can be easily learned by students. This structure should allow students to refine their common-sense beliefs into a more coherent scientific structure of the physical world.

Initially, Hestenes (1987) defined a model as "…a conceptual representation of a real thing" (p. 441) but later refined this definition by explicitly stating that models are coherent representations of the physical system studied (Hestenes, 1992).

The first attempt made to improve physics instruction using the modeling theory of instruction was researched by Halloun and Hestenes (1987) within the context of college-level instruction. During lectures, modeling theory was discussed and "modeled" via paradigm problems. When solving paradigm problems in lectures the students were guided to think in terms of the relevant information and its associated models. Two recitation sessions were taught using the deployment of the modeling pedagogy to solve additional example problems (one of these recitation sessions was given an extra two hours per week). It was demonstrated that the MDT's pre to post gain for all of the students attending the modeling lecture was greater than that of a control group of conventional students (roughly 0.42 vs. 0.23). However, the students who practiced the modeling pedagogy in recitation sessions showed even greater gains from 0.52 to 0.4 depending upon time spent on task.

A key feature in the success of the pedagogy is the structuring of physics knowledge so that it is no longer a list of equations to memorize but a coherent body of knowledge organized into a number of models. The models contain a number of distinct representations that allow the students to flexibly apply their knowledge in a variety of situations and to check internal coherence in the models developed. For example, students have both algebraic and graphical representations chunked with each developed model which can be used to predict different situations. The internal coherence of the models developed is tested whenever students demonstrate that the same prediction occurs no matter what representation utilized. While this type of lecture style deployment of the modeling theory did allow for physics knowledge to be presented in a coherent structure it did not allow for the empirical development of model representations via laboratory experimentation. During the same time frame the high school version of Modeling Pedagogy was developed. Central to the high school version was the development, revision and application of models in physical situations (Wells, 1987 and Hestenes, 1992).

This enhanced modeling method has been tested extensively. The MDT was redesigned by Hestenes, Wells and Swackhamer (1992) and renamed the Force Concept Inventory (FCI). This test was given to a number of conventional college, conventional high school and modeling method high school classes. The modeling courses showed significant gains over those from conventional classes both in high school and college. Hake (1998) compared the FCI scores for over 6,000 students based upon the degree of interactive engagement (i.e., the amount of student involvement in hands-on activities usually associated with immediate feedback from peers and instructors). Hake (1998) discovered that students in highly interactive engagement courses had normalized gain factors of about 0.7 whereas conventional courses (i.e., low interactive engagement) had normalized gain factors below 0.3. The modeling method courses in Hake's survey had normalized gains approaching 0.7. The ability of the modeling method in improving conceptual understanding as measured by the FCI continued to be demonstrated by a number of researchers (Brewe, 2002; Desbien, 2002; Vesenka et al., 1992).

The Modeling method's efficacy in problem solving has also been proven. Hestenes and Wells (1992) detail the construction of the Mechanics Baseline Test (MBT). The MBT was designed to be used by students who had prior knowledge of physics and looks like a normal quantitative physics problem-solving test. While the MBT is quantitative in nature it was also designed to test for qualitative understanding (i.e., the problems cannot be solved by simply plugging numbers into formulas) and graphical application. Hestenes and Wells found that a good posttest score on the FCI was necessary but not sufficient to produce a high score on the MBT as the correlation between the two was 0.68. Therefore, modeling students produced higher posttest MBT scores than students in conventional courses. Hake (1998) confirmed this result when he plotted his data in the same way and found a correlation of 0.91. Therefore, Hake (1998) and Hestenes and Wells (1992) determined that problem-solving ability was actually enhanced by highly interactive classes where the concepts were emphasized. These findings were replicated by several other researchers in the following years (Desbien, 2002 and Vesenka, 2002). These findings demonstrate that Modeling Instruction is a method that one can use to greatly enhance a student's conceptual understanding and their problem-solving abilities.

2.2.1 Studies of other Modeling-Based Curricula

As seen in Section 2.2 Modeling Instruction has demonstrated its efficacy through the use of paper and pencil tests focused on conceptual understanding and problem-solving ability. However, the cognitive advantages of the pedagogy have not been explored in terms of the topics reviewed in this chapter. A review of the literature for other modeling-based pedagogies such as the middle school MAR's project might determine if other research groups might have assessed additional consequences of the pedagogy such as the cognitive and metacognitive advantages.

White (1993) studied the efficacy of a course designed to develop models using the inquiry cycle and a computer microworld, known as ThinkerTools, at the sixth grade level. The entire curriculum is referred to as ThinkerTools curriculum. The curriculum was tested against a control group of naïve sixth graders and a high school physics class using a post-test transfer task consisting of 17 problems involving the concepts and principles addressed. The curriculum uses a similar approach to Palincsar and Brown's (1984) reciprocal teaching method such that initially the students were guided in a highly structured format that gradually faded away while more of the elements of the inquiry process were turned over to the students. The curriculum emphasized the development and need to translate between different representations of motion and forces. The students in the experimental group significantly outperformed both control groups on the final test. The mental models constructed by the students were explored via interviews. The interviews involved students solving out loud a series of qualitative problems. The students who did well in the ThinkerTool curriculum were able to give the correct Newtonian response (i.e., responses based on the use of an understanding of Newton's Laws of Motion) to the problems and to transfer their model to more difficult questions. However, when asked to answer far-transfer questions that covered unique situations not dealing with those specifically in the curriculum many students reverted to Aristotelian answers (i.e., answers based upon ideas such as all motion has a cause). One disadvantage of this curriculum is that while it seems to focus on models of motion and forces, which this paper is specifically researching, it does not mention how the curriculum attempts to organize the models or how the models are specifically developed. The curriculum was redesigned by White and Frederiksen (1998, 2005) to incorporate a high level of self-assessment in order to enhance metacognitive skills. They demonstrated that the metacognitive addition did enhance students' achievements on some of the assessed tasks. This curriculum design will be discussed further in Section 2.5.2.2.

Schwarz and White (2005) redesigned the ThinkerTool computer software to allow for more exposure to model development such that students would test their models by changing the computer's environment. For example, students had the ability to change the gravitational force exerted in the microworld. After testing their models the students would debate and present the models to the rest of the class. They found that the modeling assessment posttest developed by Schwarz demonstrated that the students did develop a better understanding of the nature and purpose of models but had not promoted an understanding of how models were created, evaluated and revised. No comparison between the two curriculum forms was conducted for student understanding of modeling. No differences were found between the Modeling ThinkerTools and the original ThinkerTools curriculum in the development of scientific inquiry skills and physics knowledge. It seems that the non increase might be caused by the fact that the original ThinkerTool curriculum built models implicitly so that the only new item in the modeling version was that the students were able to change some of the computer parameters to produce non-Newtonian environments. Schwarz and White (2005) did find that the modeling posttest score was highly correlated with the physics posttest thereby demonstrating a link between knowledge of modeling and the learning of science content. The link between modeling and science content shown by Schwarz and White (2005) has been seen in the Modeling Instruction research through the higher gains on the FCI for modeling vs. non modeling students. In Hake (1998) the data demonstrated that non-modeling inquiry classes do not often produce the same gain factors.

Another recent middle school modeling curriculum is called MARS (Modeling Assisted Reasoning in Science). This curriculum spans the three years of middle school and utilizes computer programs with which to build scientific models that have different interlinked representations. Raghavan, Sartosi, Schunn, and Scott (2005) demonstrate that the MARS students develop a better understanding of what models are and what they are used for after the curriculum. This finding is similar to that of Schwarz and White (2005). Lawson's Test for Scientific Reasoning was administered to both the MARS students and to a control group and the MARS students demonstrated significantly higher scores. In addition, the knowledge gain of the students was tested using a mixture of FCI, TIMMS and NAEP questions. MARS students post test scores were significantly higher than that of the control group. The MARS program is finding a similar link between modeling and increased conceptual understanding.

An elementary teacher pre-service course at San Diego State University was developed that uses inquiry activities and computer simulations that helps students construct powerful conceptual models to explain physics phenomena. The materials developed for that course are currently called Constructing Physics Understanding (CPU). Galili, Bendall and Goldberg (1993) completed a project looking at the effects the instructional units had on the students' knowledge state in the area of image formation. Galili et al. (1993) conducted interviews with students after the course using a number of tasks that included the drawing of a ray diagram and follow-up questions keyed specifically to each task. Half of the tasks used equipment that the students had used previously in the course while the other half included unfamiliar equipment. The experimenters inferred the state of the students' knowledge from their comments and their ray diagrams. They argued that the results demonstrated that the postinstruction students' knowledge about image formation represented a well-defined intermediate state of knowledge thereby showing that the students using these materials were becoming more expert-like in their knowledge state. Galili et al. (1993) concluded that since the postinstruction students' state of knowledge is well-defined but a hybrid between the preinstruction state and that of the expert state that strong restructuring was necessary to achieve an expert state. However, there was no direct comparison in this study between the CPU students' state of knowledge in this domain and that of students in more traditional classes. However, over the course of the several years students were interviewed in order to document their understanding (Goldberg and Bendall, 1995). Students were asked to explain a novel prism and concave mirror task. Similar tasks were included on the course final exam for comparison over a period of two years. It was determined that the number of major errors committed by the students when working of the tasks dropped from 79% in the 1988 interview to 24% on the 1993 exam. Goldberg and Bendall (1995) felt that these tasks "provided some evidence of the effectiveness of the approach" (p. 988). In the area of electric circuits the students were asked a question identical to one produced by McDermott (1992). McDermott (1992) found that only 10% of students in an algebra based college physics course and 15% of students in a calculus based college physics course were able to answer the questions correctly. The CPU students over the course of four semesters answered the question correctly 72% to 80% depending upon the semester.

The evidence seems to indicate that modeling based curriculums at the middle school, high school and college levels promote a greater conceptual understanding than that of conventional or other inquiry courses. Only the high school group has shown the effects of the curriculum on problem-solving ability. The problem-solving ability has been shown to be greatly enhanced over that of exiting students in conventional or other inquiry based curriculums. Most of the modeling based curriculums have not looked at the knowledge structures and problem-solving strategies developed by the students. One of the curricula discussed above did look at the knowledge states of the exiting students but did not directly compare it to that of exiting students in more traditional courses. In addition, little work seems to have been done on how students use metacognitive behaviors to further problem solving in modeling classes. The question becomes: Why might the problem-solving ability be better and why might the conceptual understanding be enhanced? In order to discover what cognitive and metacognitive traits the modeling students might be developing one must look at the research from the domain of cognitive science.

2.3 Problem-Solving Research

There have been many ways that the research community has defined problem solving over the years. Since this thesis deals with problem solving, a definition of what is meant by that term in the context of this research is needed. Polya (1968) said that problem solving was "finding a way out of a difficulty, a way around an obstacle, attaining an aim that was not immediately attainable." (p. ix). Hayes (1981) defined a problem as "whenever there is a gap between where you are now and where you want to be, and you don't know how to find a way to cross that gap, you have a problem" (p. i). Likewise, Newell and Simon (1972) described a problem in the following manner: "A person is confronted with a problem when he wants something and does not know immediately what series of actions he can perform to get it" (p.72). For most researchers in physics a problem is usually defined as the tasks listed at the end of the each chapter. These tasks are considered problems by physics researchers since there are "givens" and a question (or goal) to solve for. It is the students' job to answer the question or achieve the goal from the "givens". The problems are at the end of the chapter since the students must first obtain an initial knowledge state that will allow them to know the actions needed to solve the problems. These tasks are usually very specific, well-defined problems and may not seem to fall in line with the definitions of the psychologists above. However, for a novice in many cases the tasks at the end of the chapters are very much problems since they do not immediately know how to start nor what methods to use to reach the final goal. This seems to fall very much in line with the psychological definitions above. Of course, it is always possible that experts might consider the end of chapter tasks very trivial and not in the least problematic. In order to determine why novices experience difficulty solving problems and how we as educators might help them become better at the task, researchers designed studies that contrasted the problem-solving behavior of novices to that of experts. In the sections below I will review the pertinent problem-solving studies, focusing on those in the fields of physics and mathematics as there are many similarities between the two fields. A review of this area of the problem-solving literature will allow for a better understanding of the nature of good problem solving and why the modeling pedagogy might help students become more superior problem solvers. In addition, the review will highlight the problem-solving trajectory from novice to expert in order to determine if the problem-solving abilities of modeling students are more in line with those of an expert rather than those of a novice.

2.3.1 Problem-Solving Differences between Novices and Experts

The initial research defining the differences between experts and novices began in non-academic domains such as chess (de Groot, 1965 and Chase and Simon, 1973), taxi driving (Chase, 1982) and bridge (Charness, 1979). Researchers soon moved into academic domains such as computer programming (McKeithen, Reitman, Rueter and Hirtle, 1981) and physics (Simon and Simon, 1978). The definition of an expert in these studies was loosely defined to mean a subject or subjects who had more experience in the field or better problem-solving expertise as evidenced by grades than others. The majority of the novices were college undergraduates and the experts were graduate students or full professors. Simon and Simon (1978) studied two subjects, one expert and one novice (college undergraduate), solving physics problems via talk aloud protocols (also known as verbal protocols). The problems used were taken from a typical college physics textbook. The subjects were asked to solve the problems while saying everything they were thinking till they finished the task to their satisfaction. The subjects demonstrated several similarities: both read the problem, selected appropriate equations, and solved them after plugging in the unknown values. Simon and Simon (1978) found that the main difference between the two was the type of strategy they employed. The expert used a working forward strategy while the novice chose a working backward strategy. The difference between the two strategies dealt with where the two subjects started the problem-solving journey. The expert chose to work with variables generating a series of equations till they reached the solution (hence the term working forward) while the novice considered first and foremost the ultimate goal. The novice first defined the goal (i.e., what they were searching for) and then hunted for an equation that contained the unknown to be solved for. Hence this strategy became associated with the term working backward. Simon and Simon (1978) also discovered that there was a 4:1 difference in solution speed favoring the expert subject. The difference in the speed might have occurred since the experts seemed to use fewer equations and had shorter solution procedures than the novice subjects. In addition, at times the experts seemed to immediately recognize which equation was needed. They concluded that it seemed like the expert's solution path was guided by a type of "physical intuition" (p.337). They believed that this intuition probably meant that the expert was referring to physical principles in order to solve the problem and that this was a major reason for the success observed. However, since the problems used in this study were obtained from a standard first-year college physics textbook the expert might have simply recognized the solution to the problems. The last difference that

Simon and Simon (1978) mention is that the expert made only one metastatement per problem while the novice made five such statements on average. The metastatements were usually about planning the solution, the meaning of the equation chosen, observing errors made, and self-evaluations of progress. These metastatements will be discussed later in more detail as this observation was quite discrepant from later studies.

In the Simon and Simon (1978) study they make no reference to the expert completing an analysis of the problem situation. However, McDermott and Larkin (1978) reported that in a verbal protocol study their expert chose to complete a qualitative analysis of the problem while the novice in the study consistently avoided doing so. Larkin (1979) continued to work with expert/novice differences and found that during a talk aloud protocol study that experts paused for shorter amounts of time between retrieving equations than did the novices. Larkin (1979) attributed this finding to the possibility that the experts had the equations linked or grouped together in cognitive chunks allowing for the quick activation of other linked equations. She also said that the chunk seemed to be linked to a fundamental principle as the experts mentioned the principle when conducting the qualitative overview. Larkin (1979) hypothesized that the qualitative overview served two functions:

1. It allowed for an easy way for the expert to check the equations used against the original physical situation thus reducing errors.

2. It was a method by which the expert obtained an easy to remember overview of the problem's main features.

The imagery used to represent a problem may be crucial to the ability to reach a correct solution. McDermott and Larkin (1980) reported that experts also used diagrams representing the problem statement during their solutions. They reasoned that experts use diagrams to such a large extent because it minimizes the likelihood that they might become confused and it allows them to quickly determine if a particular solution approach is appropriate. In addition, research has shown that experts in economics use graphs as place holders of information so that it can provide cues to the next steps in a specified line of reasoning (Tabachneck, Leonardo, and Simon, 1994).

Other studies continued to find convergent results when comparing physics experts with novices such as the experts' consistent use of principles to work forward towards a solution via a qualitative analysis of the problem and the novices' tendency to work backwards via an equation that contains the unknown they are solving for (Larkin, 1981; Larkin, McDermott, Simon and Simon, 1980a, 1980b). The main focus for the novice during problem solving was the status of the unknown variable. In a number of cases the researchers designed computer models based upon skilled physics experts and novices (Larkin, 1981; verbal protocols (Larkin, 1980; Larkin, McDermott, Simon and Simon, 1980a, 1980b; Reif and Larkin, 1979). One of these models was a hierarchical planning model that worked by first reading the problem, noting the quantitative relations mentioned, planning a solution by qualitatively constructing the relationships between the major aspects of the problem, selecting quantitative equations based upon the principles generated in the qualitative construction, and then checking the solution by using a variety of different techniques (Larkin, 1980). These computer models were able to demonstrate that by following these "expert-like" procedures the expert performance empirically observed could be duplicated.

The finding that experts seem to work forward while novices work backwards was brought under question by a study of 79 subjects in a study designed by Priest and Lindsay (1992). These researchers found that novices and physics experts used a similar amount of forward and backward inference. Anderson, Greeno, Kline and Neves (1981) noted that subjects performing geometry proof generations often worked forwards as well as backwards. However, the selection of problems used in the studies may have allowed for this finding. If the problems selected are relatively easy then even novices may be able to work forward in order to solve them. Therefore, in the case of difficult problems the novices may not have the necessary knowledge to enable them to work forward. This limitation of the studies was clearly pointed out by Singh (2002) when she found that experts given very difficult problems behaved more like a novice although they still approached the problem solution in a more systematic way.

The initial study by Simon and Simon (1978) said that novices made more metastatements about planning and checking their solutions while Larkin (1979) found the opposite. There is a possibility that the reason for the differences is that the Simon and Simon (1978) study utilized regular end of chapter problems which would have been quite straightforward for the expert. Dhillon (1998) completed a verbal protocol study similar to past designs and his findings supported Larkin (1979). His findings led him to the belief that checking the solution strategy was an inherent part of the strategy for experts as they consistently checked their work and logic as they progressed towards the solution. If a novice did check their solution it was only superficially. His subjects were allowed to use a physics text as a reference if they wished. Dhillon (1998) discovered that novices consistently referred to the text for examples while the experts did not.

Schoenfeld (1985, 1987) demonstrated similar findings with math experts and college age novices. He discovered that when the novice selected an initial path they rarely deviated from it and continued down that path no matter how unsuccessful it was shown to be. The good problem solvers and math experts were goal-directed and constantly evaluated the status of their solution approach by evaluating problem-solving approaches as they were generated. These studies also implied that there seemed to be a link between better problem-solving abilities and more "expert-like" behavior.

Researchers began using younger subjects in an attempt to determine what differences existed between good and poor problem solvers. Finegold and Mass (1985) based a study on the decision that students that obtained grades higher than 90% in their high school advanced placement physics course would be considered good problem solvers while those with a grade lower than 60% would be considered poor problem solvers. The students completed five problems while thinking aloud and the good problem solvers were able to arrive at a good solution for all of the problems while only two poor problem solvers did so. Finegold and Mass (1995) found many findings similar to the initial research such as that good problem solvers translated the problem statement in more detail, planned their solutions, completed all five problems in less time, spent more time on problem translation and planning, and used physical reasoning more often. The poor problem solvers were more likely to deploy physical laws incorrectly and do little or no planning out of the solution method. No significant difference between the two groups in how often they checked their final answers was found but 3 out of 8 good problem solvers checked their solution paths while only 1 out of 7 poor problem solvers did. The difference in the numbers that checked their solutions seems to definitely favor the good problem solvers. A number of these findings were replicated in the field of electrostatics by McMillan and Swadener (1991) and college physics by Zajchowski and Martin (1993).

Santos (1995) conducted a study with ninth grade math students. Thirteen students solved math problems via think aloud methods. Their problem-solving efforts were characterized by high, medium and low level problem-solving ability based on the number of correct solutions. While this study had similar findings as the other studies in this section, Santos (1995) also demonstrated that flexibility in problem solving via the use of different representations allowed the high level students to be more successful. The lower level students who only chose a numeric representation were not able to determine the qualitative structure of the problem thereby being less likely to solve it correctly. In a study of analogy use by experts, Clement (1991) also demonstrated that experts were more flexible in reaching a solution path and usually choose to check their solutions via alternate paths. A similar good/poor problem solver design was conducted by Hegarty, Mayer and Monk (1995) with college undergraduates using inconsistent-compare problems in the domain of arithmetic. A compare problem is characterized by relation statements in which the value of one variable in the problem is defined in terms of another variable in the same problem. Inconsistent versions of a compare problem contain relational keywords (such as less) that prime an incorrect mathematical operation (subtraction rather than addition). Hegarty et al. (1995) found that successful problem solvers spent less time solving the problems, seemed to take the time to base their solution plan on a model of the situation described in the arithmetic problem, and made fewer errors. As a part of their methodology they utilized eye fixation data to determine how often and what problem elements subjects focused on. They found that the unsuccessful problem solvers referred to the problem statement more often and focused on the numeric terms of the problem while the successful problem solvers were more balanced attending not only to numeric terms but also to the problem situation as a whole. It was found that these behaviors on the part of the unsuccessful problem solvers led to their not being able to recover nor detect reversal errors (i.e., students would add when they should have subtracted or vice versa). The eye fixation data seemed to be supported by the fact that in a retrospective interview the poor problem solvers remembered more details concerning the numbers in the problem while the good problem solvers remembered more about the context of the problem situation. Savelsbergh, de Jong and Ferguson-Hessler (1996) in another good/poor problem solver design but with the addition of experts discovered a continuum of strategies with decreasing use of the methods detailed above as the subject's problem-solving skill lessened. In addition, they might have discovered a clue as to why good problem solvers are able to remember more of the context of the problem since the good problem solvers and experts in this study produced detailed elaborations of the elements in the problem. These elaborations could lead to greater understanding as shown in self-explanation effect research (see Section 2.6).

It is possible that the problem-solving differences shown above might be affected by how the expert and novice use procedural and declarative physics knowledge. Reif and Allen (1992) investigated the differences in the use of domain specific knowledge specifically in the area of acceleration. They found that while both groups invoked the concept the same number of times the novices often misapplied it especially in complex cases and did not invoke the concept components that should have been linked to the basic concept. Plus, when a novice was able to invoke the components they were unable to apply them correctly. In addition, experts were able to use supplemental knowledge that seemed to be linked to the problem concept such as forces but novices did so only moderately. A large difference was noted when it came to special cases. Experts would use case specific knowledge about the acceleration concept on familiar cases whereas the novices incorrectly applied them ignoring the case specificities. Reif and Allen

(1992) felt that the novice's concept of acceleration was not coherent and lacked the knowledge of when to apply special cases which means that the knowledge used in problem-solving situations between the two groups was very different. Actually, in the special cases the novices seemed to always apply their knowledge about special cases regardless of the specific situation. The differences in the coherence of the concept should lead to the production of more errors on the part of the novice students. This study suggests that the knowledge organization between the two groups might be different.

2.3.1.1 Summary of Problem-Solving Differences between Experts and Novices

In the review above I have detailed the commonalties and differences in the problem-solving methods used by experts and novices. In addition, I have highlighted the discrepancies between studies and attempted to explain why they might exist. A summary of these problem-solving behaviors can be seen in Table 2-1.

EXPERT BEHAVIORS	NOVICE BEHAVIORS
Typically use a working forward strategy except on more difficult problems	Typically use a working backward strategy
Performs an initial qualitative analysis of the problem situation	Usually manipulates equations discovered via equation hunting
Constructs diagrams during solution process	Rarely constructs or uses diagrams
Spends time planning approach sometimes via models of the physical situation	Rarely plans approach simply dives in
Uses fewer equations to solve the problem	Uses more equations to solve problem
Usually solve problems in less time	Usually takes more time to solve the problems
Refers to the physical principles underlying the problem	Refers to the numeric elements of the problem
Concepts more coherent and linked together	Concepts not coherent and lack applicability conditions for special cases
Fewer errors - concepts usually deployed correctly	More errors - concepts usually deployed incorrectly
Can use more than one representation to solve problems – which usually allows them to deviate to other solution paths when stuck	Usually only utilize a numeric representation to solve problems – once they become stuck rarely can free themselves
Checked solution by a variety of methods (i.e., more flexible)	Superficially check solution if at all
Rarely refer to problem statement or text	Frequently refer to problem statement and textbook (especially examples)

Table 2-1: Comparison of Expert and Novice Problem-Solving Behaviors

In a number of cases researchers mentioned that coherence of an expert's knowledge, their "physical intuition", and the equations and principles used might be chunked together and this might be associated with the observed differences. This leads one to ask what is different between the knowledge and its organization to allow for these observed differences in problem-solving behaviors. The differences between how the two groups organize their knowledge will be discussed in detail in Section 2.4. There is one significant finding that is mentioned briefly in a number of the studies described in this section - the qualitative differences in the planning, monitoring and evaluation completed by the two groups. These differences fall under the general category of metacognition. The question becomes, when you do undertake the time to make these metastatements, what does it buy the user? In order to further this finding the research conducted in this area is discussed in detail in Section 2.5.

2.3.2 Problem-Solving Strategy Training Studies

A number of research teams between the late 70's and 80's have attempted to teach problem-solving strategies to students to determine if they produced improvement in problem-solving ability. The strategies taught included the behaviors used by experts and many seem to be based on a four part strategy described by Polya (1968) – understand the problem, devise a plan, carry out the plan and finally look back. I divided the studies between ones conducted in a lab setting and ones conducted in an actual classroom setting. All of the studies mentioned were conducted with college age students.

2.3.2.1 Problem-Solving Strategy Studies Conducted in a Laboratory Setting

In the late seventies two studies, Larkin (1979) and Larkin and Reif (1979), reported that they taught an "expert-like" problem-solving strategy based upon an analysis of expert to novices verbal protocols. Five students were taught to conduct a qualitative analysis of problems based on fundamental principles and associate problem information such as equations with seven specific electricity and magnetism principles by chunking. Larkin and Reif (1979) believed that the chunking of equations would give the students an easily remembered overview to use during a qualitative analysis of the problem. Five additional students were not given any chunking or qualitative analysis trainings. All ten students then talked aloud while solving 3 direct current circuit problems. In the control group only 4 out of 5 students could solve one problem while the experimental group solved a majority of the problems (2 students solved 2 problems and 3 students solved all 3 problems). These findings demonstrated that teaching students to behave similar to experts seemed to improve their ability to solve problems. Larkin and Reif (1979) went one step further and hypothesized that that this might mean that expert knowledge was organized via coherent chunks rather than into lists of principles or equations.

Reif and Heller (1984) developed an efficient hierarchical problem-solving procedure that did not exactly reproduce the procedure shown to be used by experts but seemed to be similar but more efficient. The procedure contained three stages: problem description, solution search, and solution assessment. A number of the elements they included in the different stages were a theoretical description of the physics involved, exploratory analysis of the problem and some metacognitive processes. They thought a hierarchical problem-solving structure would be best with the top levels containing basic ideas and the bottom levels elaborating on the topmost levels. They tested the effectiveness of the procedure using 24 undergraduates divided into three conditions: experiment procedure, modified procedure developed from that used normally in textbooks and no procedure group. They assessed the solutions on post task problems for all students based upon adequacy of motion and interaction information, equations used, and correctness of the final answer. The solutions were assessed based on the written work of the students and the accompanying verbal protocol record. The experimental procedure group performed significantly better than the other two groups on all measures tested (Heller and Reif, 1984).

Lewis (1989) designed his problem-solving strategy to force novices to adopt the expert-like use of multiple representations for problems. He conducted a study using almost 100 math students to test if training students to first use a diagrammatic representation then convert it to an algebraic statement would improve their problem-solving skills. He had three experimental groups: the diagram group received problem-translation training as well as the diagrammatic integration training; the statement group received only the translation training and the control received no training. The diagram group produced significantly fewer reversal errors (i.e., selecting the inverse of the proper algebraic operation) on inconsistent-compare problems from pre to post-test and was able to transfer the newly learned skills to more complex problem situations that were similar to the training problems (i.e., near transfer tasks).

2.3.2.2 Problem-solving Strategy Studies Conducted in a Classroom Setting

Since the lab studies demonstrated a decided ability to increase the students' problem-solving ability via the training of expert-like problem-solving strategies the question now became: Could these types of problem-solving strategies that are effective in the lab be scaled up for a classroom setting with continued success even without randomly assorted subjects? Wright and Williams (1986) developed and taught a procedure called WISE to community college physics students. The WISE method had students initially identify the principles involved in the problem, draw a sketch; isolate the unknown; substitute values; and then evaluate the answer. Students who used the procedure had significantly better classroom performance than other students. In addition, these students seemed to feel that the strategy made their problem-solving process better. These results were encouraging and did not require a restructuring of the course.

At the University of Minnesota Heller, Keith, and Anderson (1992) developed a five step problem-solving approach called the Minnesota Problem-Solving Strategy for use in algebra based university physics classes employing a cognitive apprenticeship model. This strategy required a restructuring of the introductory course since the problem-solving strategy was modeled in practice sessions and lectures. The instructors in the practice sessions slowly turned the process of instruction over to the students as they become more competent. The efficacy of this type of cognitive apprenticeship pedagogical approach has been shown in numerous studies (Collins, Brown and Newman, 1989). The problem-solving strategy had the students visualize the problem, describe the physics terms, plan the solution, execute the plan, and then check and evaluate the solution. The practices made use of context rich problems that forced the students to use the developed strategy since normal novice strategies seen in some of the studies in Section 2.3.1 would not allow for success. Based on a scoring rubric developed by the group it was discovered that group solutions were consistently better than individual solutions especially in the areas of the qualitative analysis but they did not compare results to students not using the Minnesota Problem-Solving Strategy in order to determine the effects on problem-solving success. Huffman (1994, 1997) did test this idea with high school students where one subset of students was taught using the Minnesota Problem-Solving Strategy and the other using a general textbook strategy. While Huffman did find the quality of solutions better with the Minnesota group, especially problem representations, there was no difference in conceptual understanding or solution organization between groups. This study does seem to suggest that it may be very difficult to train large numbers of students in the classroom setting.

2.3.2.3 Summary of Classroom and Lab Studies

It is interesting to note that the commonality between the studies in Section 2.3.2 is that in all cases they have demonstrated that student performance in some way was improved by explicit instruction in problem-solving strategies. One major difference between these studies was that only Larkin and Reif (1979) utilized verbal protocols in order to directly observe the differences while the rest mostly focus on group administered paper and pencil tests and/or intense analyses of the students written problem solutions. Therefore, the majority of the studies and most importantly the classroom studies did not determine if the students actually internalized the strategy and were exhibiting expert-like problem-solving skills in the classroom such as working forward but could only infer that this might be what happened due to the performance on written assessments.

2.4 Schema and Knowledge Structure Research

The problem-solving studies referred to the use of principles by experts and that these principles seemed to have algebraic representations and other knowledge linked, chunked or connected in some fashion. These findings allowed researchers to postulate that the experts seemed to have developed a coherent set of knowledge that helped them perform better. Reif and Allen (1992) definitely showed that while novices and experts use the same concepts to solve problems the knowledge in the case of the novice lacks coherence. It has been suggested that the difference in problem-solving abilities might be due to the greater amount of knowledge that experts possess (de Jong and Ferguson-Hessler, 1986). However, there is some evidence for the assertion that the structure of one's knowledge may play an important role. This evidence was uncovered when Hinsley, Hayes and Simon (1977) demonstrated that competent problem solvers in algebra did indeed utilize schemas and that these schemas seemed to direct their problem-solving strategy. A schema is a mental structure that allows one to organize their knowledge. For example, in math a problem schema would consist of interrelated sets of knowledge about particular problems that unite the problems on the basis of some type of underlying feature or features. Hinsley et al. (1977) asked math students to categorize a set of algebra word problems by problem type. They found that students did indeed categorize problems into type and this categorization occurred very quickly sometimes after reading only the first statement in the problem. They went on to explore if the students used these categories to solve problems. They discovered that they did indeed utilize them to help solve problems and that the categories included information about "useful equations and diagrams and appropriate procedures for making relevant judgments" (p. 104). They also found that if the student did not recognize the problem type then they used a general problem-solving procedure which would lead to a working backward approach. Hinsley et al. (1977) concluded that this research did support the idea that people form schemas which are knowledge structures that can powerfully and flexibly represent an individual's knowledge. The card sort procedure designed by Hinsley et al. (1977) was used quite extensively to study the differences between knowledge structures constructed by novices and experts in a number of domains.

2.4.1 Differences in Knowledge Organization between Experts and Novices

In physics a classic study was conducted by Chi, Feltovich, and Glaser (1981). Chi et al. (1981) asked 8 students who had completed a semester of physics (novices) and 8 physics graduate students (experts) to categorize a set of problems taken from a standard college text based upon the similarity in their solution processes but were told not to solve the problems. They discovered that the novices sorted the problems based upon surface structure while the expert sorted them based upon deep structure (i.e., physics principles). A surface structure was considered "(a) the objects referred to in the problems (e.g., a spring, an inclined plane); (b) the literal physics terms mentioned in the problem (e.g., friction, center of mass); or (c) the physical configuration described in the problem (i.e., relations among physics objects such as a block on an incline plane)" (p. 125). Chi et al. (1981) postulated that these categories might be used by the novices and experts to access a knowledge unit that contained information they could use to solve the associated problems (i.e., a problem schema). It was suggested that the categories used by novices and experts to access their schemas would be dissimilar with novices categorizing problems via surface features while experts used principles. They then proceeded to conduct experiments that would allow them to determine the contents of these knowledge structures by asking the subjects (2 novices and 2 experts) to describe all they could about the problems in each category and how they would go about solving them. The experts associated the principles with solution procedures and applicability conditions which were similar between experts and seemed to be replicated by the Reif and Allen (1992) findings. The novices seemed not to have many (if any) explicit solution procedures associated with the surface features. These speculated structures would lead to a top down problem-solving approach by experts whereas the novices would exhibit a bottom up approach since they initially activate the specifics about the problem instead of the overreaching principle that would guide the solution approach. This would match up with the problem-solving behaviors observed in section 2.3. In addition, when Chi et al. (1981) included intermediate level students they discovered the knowledge structures were based on a mixture of principles and surface features suggesting that learning is correlated with a general shift in the structure of the knowledge organization from one based upon surface features to one based upon physical principles. This is very similar to the knowledge state finding of the CPU curriculum discussed in section 2.2.1 (Galili et al., 1993).

Chi, Glaser, and Rees (1982) continued the above studies by asking subjects to take their original sorts and to subdivide each category. They found that many of the novices' sub groupings included only one or two problems and that novices' had difficulty developing higher level categories. This suggests that the experts' develop a hierarchical knowledge structure which includes the physical objects in the problems at the lower levels whereas the novices' knowledge structure includes the physical objects in the higher level categories. In addition, it was discovered that both experts and novices use the same words to cue their solution procedures but that their reasoning was different. Chi et al. (1982) concluded that the problem-solving difficulties demonstrated by novices must be due to inadequacies in their knowledge organization. These hierarchical knowledge structures based upon principles would lead to experts conducting a breadth first search as they would need to skim over all the principles to decide upon an initial solution approach. However, novices would select a strategy based on surface features; thereby committing to a depth search as they tried every procedure connected with solving problems associated with that surface feature. Chi et al. (1981) suggest that students taught ways to reorganize their knowledge might improve their problem-solving abilities.

Chi et al.'s (1981, 1982) card sort findings were later replicated and extended by a number of researchers in the area of physics (Hardiman, Dufresne, and Mestre, 1989**;** Snyder, 2000 and Veldhuis, 1986). All three of these studies determined that some novices categorize problems based on a mixture of surface features and principles showing that the process is not as clear cut as Chi et al. (1981) may have concluded. The findings that deep structure characterizes expert categorization while surface features drive novice performance was also demonstrated in the field of mathematics (Krutetskii, 1976; Schoenfeld, 1985; Schoenfeld and Herrmann 1982; Silver, 1979 and Silver, 1981). The robustness of deep structure vs. surface feature categorization by experts and novices has been demonstrated in a number of diverse domains such as biology, computer programming, dinosaur knowledge, engineering and even aquarium usage (Chi and Koeske, 1983; Gobbo and Chi, 1986; Hmelo-Smith and Pfeffer, 2004; McKeithen, Reitman, Rueter and Hirtle, 1981; Moss, Kotovsky, and Cagan, 2006; Smith, 1990, 1992; and Weiser and Shertz, 1983).

A number of studies in physics extended the original Chi et al. (1981) study by looking at the knowledge structures developed by good problem solvers and poor problems solvers. Researchers hypothesized, based on the prior studies, that there should be a difference in knowledge structures between the two groups of students. In a forced choice study Hardiman, Dufresne, and Mestre (1989**)** found that college age novice subjects who tended to be better problem solvers also tended to categorize physics problems by principles. They were able to show that there was a correlation between problem-solving ability as demonstrated on a problem-solving task and cognitive structure based upon the percentage of problems categorized expertly. In addition, researchers soon discovered that expert or better problem solver knowledge structures were not only based more upon physics principles but that the knowledge was more coherent and that the relationship between the elements chunked or linked to the principles were explicit (de Jong, and Ferguson-Hessler, 1986; Ferguson-Hessler, and de Jong, 1987; and Savelsbergh, de Jong, Ferguson-Hessler, 1996). These findings were hinted at in the original problem-solving studies in Section 2.3.

Therefore, a series of studies looked beyond the differences in categorization and devised methods to discover the amount of cohesion in knowledge structures. Bagno and Eylon (1997) used free, cued and contextual recall tasks to probe the concepts formed by high school students of physics. They determined that the students did not have a global view of the concepts and failed to extract a knowledge structure based upon a global view. Robertson (1990) went further by investigating the type of cognitive structure high school novices developed specifically for Newton's second law during a verbal protocol study. He found that the score that students received based upon the number of connected elements normally contained in an expert knowledge structure (determined by a task analysis) correlated highly with their ability to perform transfer problems correctly. His conclusions were that if the novice is to use their knowledge structures effectively then they needed to connect the principles and concepts to each other. This same conclusion was reached recently by Sabella and Redish, (in press) when they conducted a verbal protocol study using graduate students in physics (i.e., the experts in some of the original studies) which included a structured interview. They discovered that those subjects with globally coherent knowledge structures that linked the principles together were more flexible during problem solving. In addition, subjects who had local coherence (only certain elements of the knowledge structure linked together) could not easily handle more difficult problems that required the use of several principles simultaneously.

2.4.1.1 Summary of Knowledge Structure Differences between Experts and Novices

It is very clear by the extent of the research that people do activate knowledge structures during problem solving and that not only the content but the organization of those knowledge structures correlate with success in problem solving (i.e., the more expert-like the knowledge structure the better the problem-solving ability). In addition, to be a flexible problem solver it was shown that one needs to have a global knowledge structure which links the principles together. A summary of the knowledge structure findings can be found in Table 2-2.

EXPERT KNOWLEDGE CHARACTERISTICS	NOVICE KNOWLEDGE CHARACTERISTICS
Hierarchically structured knowledge	Knowledge structure in pieces
Knowledge structure based on physics principles	Knowledge structure based upon surface features
Knowledge structures richly interconnected – global coherence	Knowledge structure mostly disconnected – local coherence
Knowledge structure links multiple representations to the principles	Knowledge structures a few usable representations
Greater amount of domain specific knowledge	Small amount of domain specific knowledge

Table 2-2: Comparison of Expert and Novice Knowledge Structures

2.4.2 Knowledge Structure Training Studies

A number of the research studies discussed in Section 2.4.1 demonstrated a link between better problem solving and more expert-like knowledge structures such that one might want to consider teaching students to categorize problems based on principles, to chunk representations of the same principle together and ignore surface features in favor of a global view. This of course assumes that there is a link between course instruction and knowledge structures since it might be possible that a more expert-like knowledge structure might be something inherent in the nature of good problem solvers and devoid of instruction. Champagne, Klopfer, Desena, and Squires (1981) demonstrated via a card sort task in a middle school class that was studying rocks and minerals that the pre to post student knowledge structures showed a marked movement towards a structure more consistent with what is considered standard in geology. This finding was replicated by several different investigators utilizing card sorts and word association tasks in math and physics (Shavelson , 1972, 1974; Shavelson and Stanton,1975; and Thro, 1978).

2.4.2.1 Knowledge Structure Studies Conducted in a Laboratory Setting

Given the findings of Shavelson (1972, 1974), Shavelson and Stanton (1975) and Thro, (1978) one might conclude that maybe it is the additional conventional problems students solve that helps produce a more expert-like knowledge structure. Sweller (1988) and Sweller and Cooper (1985) proved that this was incorrect and basically that no change occurred in a student's schema after having completed additional conventional problems. Sweller (1988) suggested that the use of means-ends analysis on the part of the novice student might produce too heavy a cognitive load that limits their ability to concentrate on the overall structure in the problem therefore showing no change in knowledge structure.

A series of studies looked at the differences in abilities produced when students studied hierarchical vs. linear materials that taught a knowledge structure directly. Eylon and Reif (1984) contrasted differences in these two types of instructional methods of acquiring knowledge in the domain of modern physics specifically its' theory and history. The hierarchical organization materials stressed a top down understanding of the knowledge which related how the concepts were linked with the general knowledge or principles in the top level and concept specifics located in the lower levels. Their argument was that this type of structure would allow students to systemically search for information. The non-hierarchical treatment basically consisted of a single level of organization consisting of the knowledge elements contained in the lower level of the hierarchical treatment. Eylon and Reif (1984) evaluated the students after they received the treatment on a number of different tasks. The students were tested to make sure they had developed the given organization and then were given free recall, cued recall and problem-solving reasoning tasks. The hierarchical group performed significantly better than the linear groups. It was also shown that these students performed better on complex tasks requiring information from several areas since the organization allowed for higher level connections between pieces of information. In addition, Eylon and Reif (1984) discovered that the hierarchical organization did not allow students to perform better on local tasks which required knowledge of isolated pieces of information. Therefore, the linkage between chunks seemed to produce greater flexibility. This finding was replicated in the area of electricity by Smith and Goodman (1984).

Taking a slightly different tack at the problem the physics education research group at the University of Massachusetts in a series of experiments attempted to determine if knowledge structures could be taught to novice students (Hardiman, Dufresne, and Mestre,1989; Mestre, Dufresne, Gerace, Hardiman, and Touger, 1993; and Dufresne, Gerace, Hardiman, Thibodeau and Mestre, 1992). They trained novices who had already completed a college physics course in mechanics to categorize problems in terms of principles using a computer program called HAT (hierarchical analysis tool). Once the students selected the principle the computer provided a set of equations that could be used to determine the solution similar to the chunking of equations observed by researchers in Section 2.3.1. They hoped that this type of tree-like hierarchical approach would allow the students to restructure their knowledge into a more “expert-like” format. They gave the students the Chi et al. (1981) card sort task pre and post experimental treatment and discovered a significant increase in performance. Looking more closely at the data they discovered that the HAT program only produced a change in those students who had not already relied on principles prior to the treatment. Having shown that the treatment with HAT could produce more "expert-like" knowledge structures they then tested to determine if the subjects showed an increase in problem-solving ability. This result was obtained when the HAT subjects demonstrated a 15% increase in problem-solving score over the control students but only when the problems used in the experimental treatment were not too difficult. When the difficulty of the experimental treatment problems increased there was no difference demonstrated between the HAT and control group. This finding lends support for Sweller’s (1988) cognitive load hypothesis. In addition, this provides a clue as to why the high school students in the Huffman (1994) study did not show an increase in problem-solving ability since the experimental group method employed the use of difficult context-rich problems. Maybe, as Sweller hypothesized, the students were unable to attend to the principles on which the problem was based due to Sweller’s (1988) cognitive load hypothesis. The context richness of the problem requires the students to use a large amount of their cognitive processing capacity in order to complete the solution process thereby being unavailable for schema acquisition based upon physics principles.

Bagno and Eylon (1997) designed a study that connected the solving of physics problems to the reorganization of high school students' knowledge structures after a year-long course. They used concept maps to aid their students in relating concepts together in a hierarchical knowledge structure in the area of electricity and magnetism. The students would actively create the concept maps while solving problems utilizing a problem-solving strategy that consisted of the following:

1) Solve – the students would solve a set of problems

2) Reflect – the students would reflect on the what principles and concepts were involved in the set of problems

3) Conceptualize – the students elaborate on the concepts and principles reflected upon in step two

4) Apply – the students apply their new knowledge and concept map to novel physics problems

5) Link – the students consistently link new concept maps developed for one problem set to the concept maps developed for previous sets of problems.

The experimental students' performance was compared to two control groups. One control received no additional training while the other did work on improving their conceptual difficulties but received no training on concept maps. Bagno and Eylon (1997) found that the experimental students performed better on all of the final tasks which included a summary of the main ideas of the domain (in order to determine the form and content of their knowledge structure), explaining the correctness of statements in electromagnetism, problem-solving ability on standard and nonstandard problems and a transfer task (asked to read a unfamiliar text and write out the main concepts and their relations). Bagno and Eylon (1997) concluded that actively constructing concept maps did indeed create a link between those concepts and problem-solving applications. Bagno, Eylon and Ganiel (2000) redesigned the materials from this study to link mechanics with electromagnetism topics and replicated the findings of the 1997 study. The study showed that one could change a student's knowledge structure after a year long course with a problem-solving strategy that was designed to help them focus on the structure and function of the connections.

Weber (2001) designed a study in the domain of mathematical proof to determine if students' inability to solve homomorphism problems was due to the lack of conceptual knowledge or the lack of strategic knowledge. He discovered that the students seemed to be missing strategic knowledge (i.e., the ability to determine when to apply their conceptual knowledge to particular problems). Weber (2001) then designed a five-step procedure that would help students to apply their conceptual knowledge to prove statements about group homomorphism. This procedure required the students to initially categorize the problems based upon four structures that were typically used to prove statements about group homomorphism problems. Weber (2001) taught the students this procedure using the cognitive apprenticeship model. It was determined that the students taught the procedure were able to construct significantly more proofs than prior to instruction.

2.4.2.2 Knowledge Structure Studies Conducted in a Classroom Setting

Van Heuvelen (1991) created a totally restructured college physics course called: Overview, case study physics (OCS). This restructuring includes a problem-solving strategy that was directly linked to principles. The structure of the class seems very similar to many of the Modeling Instruction activities described in Chapter 1. Van Heuvelen used a slightly different approach than the studies above by explicitly developing and using multiple representations, hierarchical organization charts of topic areas, problem-solving strategies and active reasoning during classes. There seem to be a number of similarities between this approach and the one used in the Eylon and Reif (1994), Bagno and Eylon (1997) and Bagno et al. (2000) studies. The physics concepts were divided into a small number of chunks each semester and developed in a hierarchical fashion. The students used and were given hierarchical organization charts to use during problem solving. The multiple representations developed included the use of graphs and motion maps. The problem-solving strategies included the evaluation of the final solution in terms of the units, sign and magnitude of the answer. This class design was taught both in precalculus and calculus based physics courses. The increase from pre to post test on the MDT (the precursor to the FCI) for both courses was significantly greater than traditionally taught classes while they also outperformed them on quantitative problem-solving measures. Compared to a conventional class the OCS students preformed better on the advanced placement physics test and on the MBT. In addition, when a group from this class and the conventional class were tested qualitatively after the course several months later the OCS students still scored higher. While this study showed dramatic results it did not attempt to discover if students did develop more expert-like knowledge structures. Although it was clear that Van Heuvelen realized the importance of a problem-solving strategy that categorized the problems based upon principles in this study we can only surmise that the problem-solving strategy might be affecting the OCS students' final performance. However, the continued performance of the OCS students after a large respite from class leads one to think that a change in knowledge structure might be one of the reasons for the performance since once the expert-like knowledge structure was in place it would be easier for them to link to it because of the interconnections rather than using a novice listing of equations.

There have been a handful of studies completed that look directly at knowledge structure development in classrooms. Keith (1993) conducted a card sort analysis of students who completed a course using the Minnesota Problem-Solving Strategy (discussed in section 2.3.2.2) which included the use of a problem-solving strategy sheet. This sheet required the students to record the physics terms associated with each problem, draw out force diagrams, and write down the equations used. Keith (1993) thought that since students using this strategy would be considering the physics terms needed to solve the problems to a greater extent they would develop a more expert-like knowledge structure. However, it is unclear how much the students would be actually referring to physics principles when describing the physics terms since the student solutions shown in Heller et al. (1992) consist mostly of forces diagrams and the relation between those forces instead of first identifying a principle such as Newton's second law. Keith (1993) tested the students' knowledge structure via a Chi et al. (1981) post course card sort and then a second sort to determine hierarchical relationships. He compared sorts between users and nonusers. The two groups were distinguished based upon test performance (i.e., those that consistently used the strategy (Users) vs. those that did not (Non-Users)). Based upon the card sort Keith (1993) found that the only significant differences in knowledge structure were at the subordinate level, what Sabella and Redish (in press) might consider local coherence. Therefore, the students using the strategy demonstrated a knowledge structure similar to physics experts at the lower levels of their hierarchical structure (i.e. contained solution procedures) but the upper levels were not organized based upon fundamental principles. Unfortunately, this study detracts a bit from the others since it seems to cast a shadow on the ability to teach students by a method that allows them to obtain a more expert-like knowledge organization. However, given that the problem-solving strategy sheet does not specifically have students start from physics principles in a way similar to the laboratory based studies one could say that this lack might have caused the knowledge structures developed to be less expert-like than if they had started the method with a principle selection.

Leonard, Dufresne, and Mestre (1996) implemented a strategy in a calculus based physics class that they hoped would help students develop a more expert-like knowledge structure. The strategy included a qualitative description of the problem starting with selecting the principle involved in the solution, justification for selecting the principle, and a procedure to use the principle to determine a solution. The students were encouraged but not required to use the strategy when solving homework problems; however, they were required to complete a strategy writing task on exams. The researchers determined that the principle trained students performed better than students from a traditional class on a forced choice categorization task designed by Hardiman et al. (1989). Therefore, it would seem that forcing students to initially categorize problems based upon principles produces students with a more expert-like knowledge structure than conventionally trained students. The deficient strategy writing task samples (only taken from the experimental class) displayed a focus on surface features; however, two-thirds of the class completed good sample tasks based on principles. Therefore, the strategy writing task seems to back up the more expert-like knowledge structures obtained by the experimental class. They also tested the students' knowledge by a free recall task where the students from both classes were asked to identify the important ideas used to solve problems. The two groups (experimental vs. traditional) identified Newton's three laws of motion with the same frequency but the strategy class identified the four remaining principles (conservation of energy, conservation of momentum, angular momentum, and work-energy theorem) at a higher frequency. This study seems to contradict the Keith (1993) study since this method used an even less intensive strategy as there were no problem-solving sessions incorporated in the course. However, this also seems to support the idea that one of the reasons for the Keith (1993) failure might have been because the strategy did not highlight the principles used to solve the problem. It is somewhat disappointing that the Leonard et al. (1996) study did not attempt to appraise the differences between the two classes in the area of conceptual understanding nor problem-solving practices, thereby correlating them to the knowledge structure obtained on the part of the student. However, the lab-based studies performed by this group have already shown that there is a correlation between the two so therefore they may not have felt it was necessary.

Alan Schoenfeld (1985) developed a mathematics based problem-solving course that incorporated the use of deep structure and metacognitive skills while solving problems. The students of the course demonstrated a significant increase in problem-solving ability compared to a control group (which showed no change pre to post). Schoenfeld demonstrated via a card sort and a statistical cluster analysis that the knowledge structures of his students' pre to post course were highly correlated to that of expert mathematician. The correlation between the two knowledge structures pre course was 0.54 whereas the post course value was 0.72 even though the course did not specifically address problem perception although the techniques used did increase the students' attention to the problem structure since they were looking for examples and examining goals (Schoenfeld, 1985; Schoenfeld and Herrmann, 1982). This course was a semester long intervention that was extremely intensive and did show that the students had a large increase in problem-solving ability but did not correlate the obtained knowledge structure to that ability.

Chabay and Sherwood (2002) developed a Modern Mechanics course called *Matter and Interactions I: Modern Mechanics*. One of the goals of the course is to involve the students in attempts to predict and explain physical phenomena using the fundamental principles of physics thereby stressing the coherence of the conceptual structure of the physics (Chabay and Sherwood, 2006). This text is designed around the application of fundamental physics principles (i.e., momentum principle ($\Delta p=F_{net}\Delta t$), angular momentum, etc). The students construct simple models of physical systems and the number of problem sets has been decreased since they are more realistic, complicated and encourage the initial categorization of a fundamental principle for ease of solution. The course instructors model the procedure of solving problems based upon fundamental principles in lecture and this is followed up in recitation sessions. In addition, the students use a 3-D computer language called VPython. The VPython programming language allows students to simulate physicals systems by the construction of programs based upon symbolic vector algebra and to visualize external vector representations of the system in 3D. In order to program a working model the students learn they must start with a fundamental physics principle and program forward from there. This course seems to implicitly teach the students a coherent knowledge structure. The efficacy of this approach has been reported by Kohlmyer (2005). Kohlmyer found some interesting qualitative differences between Matter and Interactions (M&I) and traditional students. It was determined that during talk aloud problem solutions M&I students started their problem-solving paths by invoking a fundamental principle thereby demonstrating a more expert-like problem-solving behavior. On the other hand, the traditional students emphasized equations such as F=ma and special case formulas during their talk aloud problem solutions. This finding implies that the students may be developing a more expert-like knowledge structure and completing an initial breadth search of that structure.

A companion text developed by Chabay and Sherwood (2002) called *Matters and Interactions II: Electricity and Magnetism* uses the emphasis on fundamental principles to teach students during a second semester physics course. Chabay and Sherwood identified the principles that were fundamentally important in order to develop the material in a coherent fashion. They completed a complex problem-solving study by including three standard problems on the E&M final exams in one traditional class and one class using the M&I sequence. There was no significant difference between the numbers of students who completed two of the problems correctly. However, on the third and most complex problem the M&I students' performance was four times higher than that of the traditional students. The efficacy of the M&I electricity and magnetism sequence was also tested by Thacker, Ganiel and Boys (1999). Thacker et al. (1999) used both a questionnaire to probe traditional and M&I students understanding of the transients in dc circuits and an interview to test for effectiveness. The M&I group performed better on the understanding of the transient phenomena and were able to give valid explanations even of situations unfamiliar to them while the traditional groups had a tendency to rely on algebraic manipulation as a means of explaining the situations. Engelhardt and Beichner (2004) discovered that the M&I students outperformed traditional students on a test for understanding of dc circuits they designed called DIRECT. Ding, Chabay, Sherwood and Beichner (2006) used the BEMA (Brief Electricity and Magnetism Assessment developed by Sherwood and Chabay in 1997) to analyze the effectiveness of traditional and M&I electricity and magnetism courses in a longitudinal study. It was discovered that M&I students who had scored a B in the original course obtained the same score on BEMA as that of traditional students who had scored an A in their original course. This effect was observed over the course of five semesters. When using the BEMA as a pre and post-test assessment for several sections of traditional and M&I courses it was determined that the M&I students scored significantly higher producing twice the gain as traditional E&M students. It is possible that the development of a more expert-like knowledge structure is allowing the M&I students to demonstrate this enhanced performance.

2.4.2.3 Summary of Classroom and Lab Knowledge Structure Studies

The studies both in and out of the classroom showed that one can teach students to reorganize and/or develop a hierarchical knowledge structure producing varying rates of success depending on the methods used. A correlation between problem solving and knowledge structure was suggested in the findings of a majority of the lab based studies and in a number of the classroom studies. However, none of the studies reviewed attempted to quantify card sorts in order to correlate the "expert-likeness" of the knowledge structures with problem-solving ability. It is possible that the hierarchical knowledge structure helps one become a better problem solver because it consists of a basic set of principles that students can apply with the representations for these principles chunked together, so that the problem-solving process should become easier. It should be much easier for students to choose between a handful of principles vs. an unwieldy and unending list of equations attached to surface features. It is no wonder that the novice student has such difficulty with problem solving since their inexpert-like structure requires them to select a "correct" process from a number of choices since they would normally be searching through a large equation list while the expert only must select from a handful.

2.5 Metacognition Research

So far this paper has described a number of methods that would allow students to become better problem solvers via changes in knowledge organization and problem-solving methods. The question becomes how one manages the process from beginning to end in an efficient manner so that a correct solution can be reached. Schoenfeld (1992), in a review of metacognition and mathematics, said "it's not what you know; it's how, when and whether you use it" (p. 355). Studies have demonstrated that most students do not develop proficient control strategies and their ability to solve problems is lessened (Schoenfeld, 1985).

But the question becomes just what is metacognition? Metacognition was defined by Flavell (1976) as the following:

> "Metacognition refers to one's knowledge concerning one's own cognitive processes or anything related to then, e.g. the learning-relevant properties of information or data. For example, I am engaging in metacognition….if I notice that I am having more trouble learning A then B; if it strikes me that I should double-check C before accepting it as a fact; if it occurs to me that I had better

scrutinize each and every alternative in a multiple-choice type task before deciding which is the best one…Metacognition refers, among other things, to the active monitoring and consequent regulation and orchestration of those processes in relation to the cognitive objects or data on which they bear, usually in the service of some concrete (problem solving) goal or objective." (p. 232)

Flavell (1979) encouraged the training in metacognitive skills as he felt that training in these specific skills should allow not only for improved learning to occur but also a greater amount. In addition the habit of using metacognitive skills should be useful in a number of fields and not only in the field in which they were initially trained since they are general rather than specific.

In order to study metacognitive processes one must have a framework on which to identify and categorize metacognitive aspects of problem solving. Paris and Winograd (1990) described metacognition in math education as self-management reflected "in the plans that learners make before tackling a task, in the adjustments they make as they work, and in the revisions they make afterwards. " (p. 18). Silver (1989) described these self-management processes as planning, monitoring and evaluation. This study will use Silver's (1989) description of these processes as the framework on which to identify and categorize the metacognitive processes and behaviors observed.

2.5.1 Differences in Metacognition between Experts and Novice

If the hypothesis is that metacognition is necessary for being a successful problem solver then one would expect to observe differences in the metacognitive abilities between experts and novices or between good and poor problem solvers as has been seen in the case of knowledge structures and problem-solving strategies. In section 2.3.1, Simon and Simon (1978) found that novices made more metastatements than experts. However, a limitation of this study was that the problems were very simple for the experts such that they probably had to do little planning, monitoring or evaluating of the problem-solving process. Other problem-solving studies in physics did discover that experts and successful problem solvers made a qualitative analysis of the problem or underwent reflective thinking about the problem which under the chosen framework could have been coded as metacognitive statements in the area of planning or monitoring (Champagne, Klopfer, and Anderson, 1980; McDermott and Larkin, 1978; and Larkin, 1979). Other studies in physics, math, Lisp programming and biology found that during verbal protocols experts and good problem solvers seemed to be constantly evaluating their progress towards a solution and that these subjects demonstrated improved task performance (Dhillon, 1998; Pirolli and Bielaczyc, 1989; Schoenfeld, 1983, 1985, 1987; Smith and Good, 1984; Veenman, and Verheij, 2003; Zhang, Wu, Fretz, Krajcik, Marx, Davis, and Soloway, 2002). Therefore, it seems that there might be a connection between metacognitive skill usage, problem-solving ability and knowledge organization.

It would seem that the monitoring of one's comprehension during problem solving is an important behavior if one is to be a successful problem solver. Evidence of metacognition differences between good and poor problem solvers was discovered in studies conducted about self-explanations. Chi, Bassok, Lewis, Reimann, and Glaser (1989) analyzed self-explanations made by good and poor problem solvers in physics as they studied worked out examples. The good problem solvers produced self-explanations that were guided by active and accurate monitoring of their comprehension of the material. Chi et al. (1989) found that detection of comprehension failures did initiate explanations for both good and poor problem solvers. Eighty-five percent of the good problem-solvers' detections of comprehension failures were followed by explanations describing their understanding; whereas, only sixty percent of poor problem solvers' followed comprehension failures with explanations. When the poor problem solver did produce an explanation following the detection of a comprehension failure they were usually about quantitative expressions in the problem while the good problem solvers' explanations were split between quantitative expressions and ones explaining the physics principles and concepts. This same effect was discovered by Ferguson-Hessler and de Jong (1990) when they discovered that when studying a physics text poor problem solvers failed to detect and therefore correct their comprehension failures. Therefore, these findings suggest that experts and good problem solvers might use metacognitive skills differently and more often than poor problem solvers or that the difference in use of these metacognitive skills might be due to the greater degree of conceptual understanding on their part.

2.5.2 Metacognition Training Studies

The earlier studies dealing with the training of metacognitive skills occurred in the area of comprehension monitoring in reading and the use of these strategies by "retarded students". It was found that indeed metacognitive training of "retarded students" allowed them to assess and check their readiness to complete serial recall tasks. (Brown, Campione, and Barclay, 1978 and Lawson and Fueloep, 1980). In addition, Brown et al. (1978) demonstrated that the individuals in their study were able to use the strategies a year later and to transfer them to other recall tasks thereby, showing the general nature of the skills taught. In the area of reading August, Flavell and Clift (1984) found that skilled readers demonstrated greater amounts of comprehension monitoring and when monitoring comprehension skills were taught to fifth graders, they improved their reading ability. These early studies showed that metacognitive behaviors can be very important skills. It would follow that these same skills could be important contributors to improving problem-solving abilities in students. In the preceding section a number of the problem-solving studies discussed had a metacognitive component built into the strategies taught to the student. Therefore, the significant findings in these studies may be due in some part to the metacognitive behaviors they engendered on the part of the subjects. However, none of the studies tested to see to what extent the subjects utilized the different behaviors taught by the strategy – both cognitive and metacognitive for the most part.

2.5.2.1 Metacognition Studies Conducted in a Laboratory Setting

There are not many metacognition studies that occurred in a laboratory setting that fit our framework for problem solving except for studies in the area of self-explanations and then only in the area of Lisp programming. Bielaczyc, Pirolli, and Brown (1995) identified a set of metacognitive strategies used by high performers in previous studies which included monitoring their comprehension, learning activities and clarifying and addressing their comprehension failures. Bielaczyc et al. (1995) divided their participants into two groups. Both groups received Lisp training but the control group received none of the strategy training developed for the experimental group. The two groups were balanced based upon Lisp programming performance levels. The experimental group became familiar with asking *why* questions, summarizing the main ideas and were given self-monitoring questions (such as do I understand this and what is the purpose of such and such). Therefore, the experimental group was trained in techniques that identified and elaborated on the relations between the main ideas in the text, looked at the examples in order to determine the form of the code and then explicitly connected the concepts between the text and the examples studied. This seems very similar to some of the techniques used by studies attempting to develop more “expert-like” knowledge structures. After a verbal protocol analysis of the pre and post programming lessons it was determined that the experimental group performed significantly better than the control group by producing fewer errors, making more monitoring comprehension statements and clarifying a greater number of comprehension failures. This study did not monitor metacognitive abilities in planning or evaluation directly.

2.5.2.2 Metacognition Studies Conducted in a Classroom Setting

Some of the initial studies to train students in metacognition behaviors in a classroom setting occurred in the area of reading; such as in Palincsar and Brown (1984). Palincsar and Brown (1984) taught students via a cognitive apprenticeship model four strategies designed to foster and monitor reading comprehension. The four strategies were predicting, questioning, clarify and summarizing. The student demonstrated marked improvement in their reading abilities by the end of the intervention. Was the success demonstrated due to the cognitive apprenticeship pedagogy which allowed for the emulation of the behaviors modeled by the instructors or the new metacognitive skills the students developed or a combination? Could the large success have been caused by the pedagogical usage allowing for the students to internalize the metacognitive skills to a much greater extent than other past interventions? Salomon, Globerson and Guterman (1989) utilized a computer program to teach metacognitive skills to a seventh grade experimental group whereas the control group only received non-strategic advice (such as read more carefully). They were able to show that the significant difference between the two groups could be accounted for solely by the metacognitive training.

The domain of mathematics has also been very active in the metacognitive training arena. Most importantly, Schoenfeld’s (1985) college problem-solving course contained a very large metacognitive aspect. Schoenfeld explicitly role modeled the decision making processes during problem solving. The discussions always started off with the question: “What do you think we should do?” (p.221) thereby initiating a planning sequence. Multiple ideas of initial problem-solving starting points were asked for and the class planned which path to take. Once the class reached a solution they always evaluated the final solution. As the students solved problems on their own Schoenfeld was always asking them “Why are you doing that?” and “How does it help you?” (p.221). Questions such as these ask the students to monitor their comprehensions and explicitly evaluate their progress towards the solution. Schoenfeld analyzed video taped student solutions pre and post course and discovered that the students were using a greater number of metacognitive skills thereby demonstrating more control of the solution path and performed more expertly post-course. Schoenfeld (1985) also demonstrated that the students in the course performed significantly better than a control group on problem-solving tasks. This analysis is limited as there was no analysis of error revision and correction via verbal protocols between the control group and those students taking the problem-solving course.

A number of studies in middle school math were completed by Mevarech and her associates. Mevarech and Kramarski (1997) developed a strategy, called IMPROVE, and showed that students using this strategy outperformed a control group. IMPROVE stands for "introducing new concepts, metacognitive questionings, practicing, reviewing and reducing difficulties, obtaining mastery, verification, and enrichment" (p.369). The strategy taught students possible strategies to solve the problem, to ask metacognitive questions dealing with comprehension and to make connections between this problem and past problems (in this way they might learn to categorize the problem types). The metacognitive questions were placed on strategy cards for ease of usage by students and were designed to help students become aware of self-regulation by planning the solution, monitoring the progress and allocating resources. The IMPROVE groups significantly outperformed the control group in all areas except that the low-achieving students did not demonstrate any increases in mathematical reasoning. In an additional study using a metacognitive cooperative problem-solving experimental group that used IMPROVE, a cooperative problem-solving experimental group and a control group showed that the metacognitive group outperformed the other two on all measures while the cooperative group outperformed the control (Mevarech, 1999). This same study demonstrated that the low ability students performed best under the metacognitive strategy and that the metacognitive strategy group was able to solve significantly more complex transfer problems than the other two groups.

The only high school metacognitive study in the domain of physics was conducted by Neto and Valente (1997). In this study, the authors taught one high school class a metacognitive strategy for solving problems while the other was a conventional physics class. The metacognitive group also studied more difficult problems thus possibly allowing for the development of enhanced problem-solving processes due to the complexity of the problems and not solely due to the metacognitive strategy. The metacognitive group did outperform the traditional class on both qualitative and quantitative problem sets. Neto and Valente (1997) did not look directly at the metacognitive abilities used via think aloud protocols; but, only obtained a sense of student usage of metacognitive strategies by a questionnaire administered about the usage of such techniques.

White and Frederiksen (1998, 2005) developed an inquiry curriculum called ThinkerTools which explicitly taught self–assessment in the form of self-monitoring and evaluation and utilized the ThinkerTool microworld. The ThinkerTool students reflected at the end of each unit on their processes and final product while the control group completed the same ThinkerTool curriculum but did not reflect on their own processes. The experimental group's reflective assessment gave them the advantage over the control group in areas of scientific inquiry and science knowledge as it related to the models developed but there was no difference between the two groups in applied physics problems (these were questions similar to the FCI). At first, this may seem to be a discrepant finding. However, since metacognitive strategies were only modeled in the inquiry and model development phases with the class it is possible that they did not know how to transfer these skills to a qualitative problem-solving domain.

2.5.2.3 Summary of Classroom and Lab Metacognition Studies

The studies in metacognition have demonstrated an improvement in performance in a number of areas. However, in a number of these studies the metacognitive strategy is confounded because it is taught along with other improvements to the curriculum such as general self-explanations, general problem-solving strategies, inquiry concept development, and more complex problem analysis. However, this is typically what happens in research dealing with classroom studies and can be said about many of the studies reviewed in this thesis. Through the use of verbal protocols the Bielaczyc et al. (1995) and Schoenfeld (1985) studies demonstrate that the experimental group used more self-regulation post-treatment. In addition, Bielaczyc, wt al (1995) show that the number of errors decline; however, they do not directly connect this decline with metacognitive abilities.

In addition, if one looks back at the studies in problem-solving strategies there are a number of them that also include some form of metacognitive processes. Were the metacognitive processes responsible for the increase in post-test performance in those cases? Or is it more likely that problem-solving heuristics included in the strategies, the metacognitive strategies and the development of a more expert-like cognitive structure overlap with each other?

2.6 Studies Linking Metacognition, Problem-Solving Behaviors and Knowledge Structure – the Case of the Self-Explanation Effect

A number of studies have shown that the number of self-explanations students produce correlates with problem-solving ability (Bielaczyc, Pirolli, and Brown, 1995; Chi, Bassok, Lewis, Riemann, and Glaser, 1989; King, 1992; Nathan, Mertz, and Ryan, 1994; Neuman, Leibowitz, and Schwarz, 2000; Pirolli, and Recker, 1994; Renkl, 1997; Siegler, 1995, 2000; Webb, 1989)**.** By self-explaining text examples the students should be connecting the main ideas together therefore producing more interconnected chunks which should aid in recall as shown in expert/novice studies seen in Sections 2.3.1 and 2.4.1. As shown in the review of metacognition studies the self-regulation skills of planning, monitoring and evaluating their comprehension and strategy use should help students determine if they are achieving a good understanding of the materials and enabling them to catch any errors produced. This is similar to the idea Chi (2000) postulated - that one needs to be actively involved in order to acquire new knowledge and to reorganize one's knowledge structure. She postulates that students develop self-explanations when they find a discrepancy between the text and their own mental model. Therefore, it would make sense that a greater number of self-monitoring questions would lead to finding more of these discrepancies thus leading to a greater reorganization of one's knowledge structure. Therefore, self-explaining could be one method by which one becomes an expert and develops an expert knowledge structure. Studies in this area that prompt students to self-explain show the same correlation between the number of self-explanations and problem-solving abilities; therefore, it seems that having students self-explain might allow for the development of a more expert-like knowledge structure.

Therefore, in order to answer a portion of the questions posed at the end of the last section, a review of the studies linking metacognition to the production of more expert-like knowledge structures via self-explanations is required. Chi, M.T.H., de Leeuw, N., Chiu, M and LaVancer, C. (1994) prompted an experimental group of eighth graders to self–explain a textbook chapter on the circulatory system while a control group was allowed to read the material twice. The experimental group was asked to self-explain after each sentence of text (i.e., explain the meaning of each sentence). The pre to post test comprehension questions showed that the experimental group had a 32% pre to post test gain while the control only had a 22% gain. However, much more impressive was the fact that if one looks only at the more difficult questions the gain is 22.6% vs. 12.5% demonstrating that the prompted group understood the material more deeply as these questions required the production of knowledge inferences. In addition, when Chi et al. (1994) split the prompted group up into hi-explainers vs. lo-explainers it was found that the hi-explainers generated a greater gain. In addition, the students were allowed to use the text book on the test. The hi-explainers referred to the textbook examples 2 times vs. 11 times by the lo-explainers. This is very similar to research on the problem-solving behaviors between expert and novice subjects where the experts referred back to materials to a lesser amount. This suggests that the hi-explainers are more expert-like thereby having constructed a more coherent expert-like knowledge structure. Chi et al. (1994) compared the mental models of the circulatory system produced by the prompted and control groups. The prompted group produced a correct model 57% of the time vs. 22% of the time for the control group. In addition, within the prompted group the hi explainers all attained the most accurate model possible while only 1 out of four of the lo-explainers did so. Finally, it was determined that at least 30% of the self-explanations produced by the students actually integrated new information with their older existing knowledge. However, one-fourth of all self-explanations were incorrect but they still allowed for the production of more expert-like mental models. It is possible that the integration of incorrect information could allow the students to experience conflict when comparing it to more correct information thus leading to a more correct mental model. These findings support the idea that self explanations allow students to reorganize their knowledge structures towards a more "expert-like" structure.

Since self-explanations have been shown to have a core component of metacognition involved in their usage Aleven and Koedinger (2002) designed an experiment using the computer-based cognitive tutor for geometry. The only self-explanation that was asked of the students was to name the principle that would be used to solve the problem. Aleven and Koedinger (2002) believe that this would allow them to be more metacognitive and in the framework initially discussed this procedure would allow the students to start planning the solution method. However, it is also very similar to the studies done by Hardiman et al. (1989) that showed problem-solving improvement when asking students to initially categorize physics problems based upon principles. Using this procedure Aleven and Koedinger (2002) were able to demonstrate that the explanation group significantly outperformed the non-explanation group on the post–task. However, the experimental group also spent a significantly larger amount of time completing the training. Therefore, they performed a second experiment controlling for time spent. The students were advanced out of each level when they either reached mastery or met the time limit. In this study the gain on the problem-solving task was only marginally significant for the experimental group. However, their performance on the questions requiring a deeper understanding of the concepts was significantly better than the control group while the control group performed better on those questions needing only shallow knowledge. This is very similar to Chi et al.'s (1994) finding about self-explanation and the circulatory system. Finally, when looking at errors the experimental group produced fewer errors by commission. The findings that the number of errors decline are similar to the earlier studies on expert/novice differences seen in several sections of this thesis.

Van Lehn and Jones (1993) analyzed the Chi et al. (1989) data to determine if the self-explanation effects could have been caused by students uncovering gaps in their knowledge and then filling those gaps. Van Lehn and Jones analyzed all of the errors produced by both the high self-explainers and the low self-explainers. The types of errors made by both groups were classified in order to determine which errors were reduced by self-explanation. They discovered that only gap errors were significantly different for the two groups. A gap error was caused by a lack of knowledge concerning a physics principle or concept. For example, some students were unaware that an inanimate object such as a table applies a force of any object interacting with it thereby demonstrating a gap in their knowledge. These gap errors accounted for most of the error rates between the two groups. They also analyzed if schema selection and analogical problem solving could be producing the self-explanation effect. However, the differences in errors due to incorrect schema selection could not account for the difference in scores between the two groups. The idea that the high self-explainers might be producing greater number of inferences also was disproved. Therefore, it seems that self-explanations might be allowing one to reorganize the knowledge structure as postulated earlier and over time producing an expert-like knowledge structure.

2.7 Central Features of Modeling Instruction

The evidence seen in Section 2.2 affirms that a science curriculum based on constructing models produces students that have greater abilities in the areas of problem solving, conceptual understanding, and scientific reasoning. However, in order to continue to improve on the Modeling Instruction pedagogy I believe that one needs to understand the cognitive and metacognitive skills that are developed by modeling students and how research in cognitive science informs one about these processes of change. In this section I will review the central tenets of Modeling Instruction and connect them to the pertinent literature reviewed in this chapter.

2.7.1 Model Development

During model development the students empirically obtain data to develop a model of a physical system via an experiment which they have designed. They use the data to produce several representations of the model (see figure 2-1): verbal, diagrammatic, graphical and algebraic.

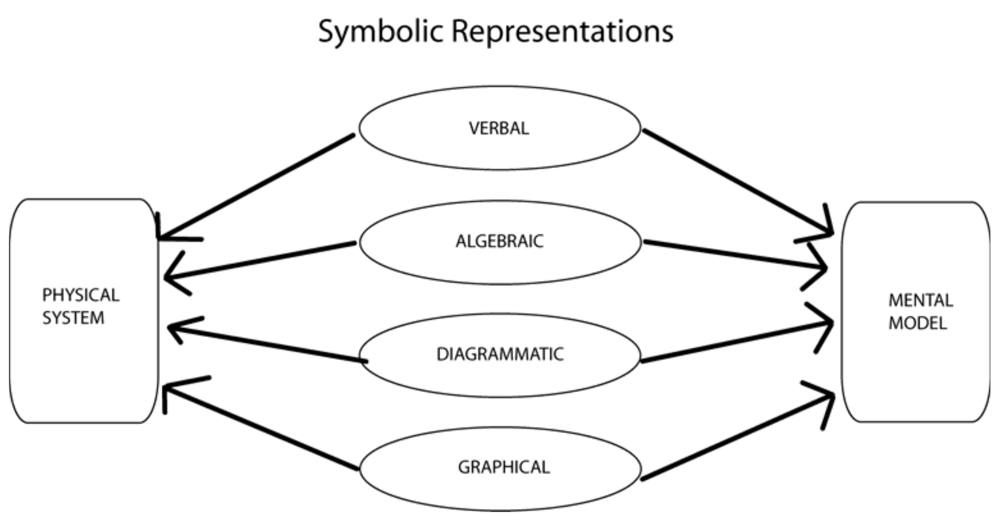


FIGURE 2-1: Explicit Representations produced between the physical system and the mental model (adapted from presentation by Swackhamer and Dukerich)

The students present their model representations to the rest of the class and justify the conclusions. The class as a whole arrives at a consensus concerning the representations generated. Although there are a number of beneficial processes occurring in this period of the class, I will focus on those dealing with the main sections of this literature review. The students are organizing their understanding of the concept in question by chunking together all of the needed representations. This is an activity that infrequently occurs in a conventional class. The students are constructing a model that has local coherence since all the representations are required to arrive at the same conclusion. As they move through the course they develop a series of models (See Figure 1-2) which have local coherence. For example, the model of constant acceleration is on the same level as Newton's second law and each branches off toward its own representations. However, global coherence might be developed if the dependence on acceleration in Newton's second law allows the students to develop a linkage to the constant acceleration model developed earlier. This process is very similar to studies reviewed in Section 2.4.2 that dealt with the teaching of a knowledge structure that is hierarchical vs. linear such as in Eylon and Reif (1984), Bagno and Eylon (1997) and Bagno et al. (2000). Therefore, one would expect to reap similar rewards as shown in these studies such as increased problem solving and conceptual understanding. While the Modeling research has demonstrated a correlation between these two they have not linked either to the development of a knowledge structure by modeling students.

The model development stage also scaffolds the students in the use of metacognitive skills since the students when justifying their work are continually asked the following questions: "How do you know that?", "Why do you think that?" and "Does that make sense?". These questions are similar to the ones used by Schoenfeld (1985) in his problem-solving course which has been proven to increase the metacognitive skills of his students. In addition, this is very similar to the self-explanation prompts used by Chi et al. (1989) during their self-explanation studies. As demonstrated by Bielaczyc et al. (1995) one would expect modeling students to self-monitor their understanding and be able to discover comprehension failures which should lead to error corrections and revisions. Finally, during this development the students are making connections between the current model and the past ones which is similar to what occurs when using the IMPROVE strategy (Mevarech, 1999; and Mevarech and Kramarski, 1997).

2.7.2 Model Deployment

During the modeling deployment stage the students deploy the model developed in the paradigm lab, the initial lab conducted in each modeling cycle, to new contexts so that the students can abstract the model thereby allowing them to use it in other situations. This is done by deployment labs when they test their models to determine if they are predictive and by problem sets similar to regular physics textbook problems except that the initial deployment problems ask the students to solve them using the various representations of the model. During the problem-solving deployments the teacher models and scaffolds the student in the use of a problem-solving strategy based upon the physics models. Before every problem the students are asked to first select a physics model that one can deploy to solve the problem thereby reinforcing the idea of an initial breadth search across all physics models which is similar to method used in Aleven and Koedinger (2002). From this physics model they have several representations with which to solve the problem. The requirement that the students first select a principle is very similar to the strategies used in the research attempting to produce more "expert-like" or coherent knowledge structure in students such as those completed by Chabay and Sherwood (2006), Hardiman et al. (1989), Leonard et al. (1996), Larkin and Reif (1979) and Van Heuvelen (1991). In addition, the use of multiple representations allows the students to be more flexible when one strategy does not work, as shown by Lewis (1989). As students are working through a problem solution they are always reminded that after planning they must monitor their comprehension and evaluate the final solution. During whiteboard presentations the students get to see the various ways that one can solve a particular problem, reinforcing the fact that there is not only one way to skin a cat but also justifying their answers as the students are always asked: "Does that make sense to you?" and "How do you know that?". Therefore, this should increase the metacognitive skills of planning, monitoring, and evaluating as shown by Schoenfeld (1985). In conclusion, the students are asked to constantly reflect and explain to themselves what they know and how they know it in all areas of the class: notebook reflections/homework, whiteboard presentations, and tests and quizzes. This type of self-explanation prompting has been demonstrated to help students construct more "expert-like" knowledge structures (Chi et al., 1994). In addition, Van Lehn and Jones (1999) found that it is likely that the knowledge structure is changed because self-explanations discover gaps in the structure which need to be filled. This is modeled in the modeling classroom by demanding that the models have local coherence such that the minute that a discrepancy in predictions is observed one goes back and refines the model.

In addition, several researchers (Aleven and Koedinger, 2002; Bielaczyc et al., 1995; Hegarty et al., 1995; and Lewis, 1989) demonstrated that the use of "expert-like" skills and behaviors seems to allow one to produce fewer errors and also possibly catch those errors more frequently. There has been little research into this error production and its revision. However, one might expect that modeling students would produce fewer physics errors.

2.7.3 Skill Expectations

This literature review hints at a number of possible skills that the modeling students could be developing and using through the course of a school year that non-modeling students would not be able to develop easily, as the modeling course activities implicitly and explicitly contain a number of the strategies used in the studies reviewed as shown above. Reviewing section 2.7.1 and 2.7.2 one would expect modeling students to have a more "expert-like" knowledge structure and exhibit more "expert-like" problem-solving behaviors such that when solving problems students would:

- Use multiple representations to solve problems (such as graphical methods vs. algebraic methods)
- Identify the method of solution via models instead of equations
- Complete a breadth search of knowledge structure instead of a depth search
- Use metacognitive skills continuously (setting goals, monitoring, and evaluating)
- Produce fewer physics errors

This thesis brings together the research in all these areas reviewed to further the understanding of what problem-solving behaviors, knowledge structures, and metacognitive skills are developed by modeling students vs. non-modeling students. One can see that there are a number of holes in the cognitive and metacognitive underpinnings of Modeling instruction that need to be filled. There have been no classroom studies correlating a student's knowledge structure to their problem-solving ability (only lab based studies) and no studies correlating conceptual understanding as shown on the FCI to one's knowledge structure. For that matter there have been no studies that show that the modeling students even internalize a more "expert-like" knowledge structure. There have been no studies conducted to observe and compare the specific problem-solving characteristics and metacognitive skills utilized between modeling and non-modeling students. The research described in the following chapters expands on the present findings of the Modeling Instruction method to include experiments designed to answer the following questions:

1) Do the modeling students have a more "expert-like" knowledge structure?
2) How are the conceptual understanding, problem-solving abilities and knowledge structures of modeling students correlated?
3) How will the problem-solving strategies and metacognitive strategies of the modeling students differ from that of the non-modeling students?

Chapter 3
Study 1: The Correlation between Knowledge Organization and Performance

3.1 Purpose

The purpose of Study 1 was to compare first and second-year modeling and non-modeling participants' conceptual and problem-solving abilities in the area of mechanics and energy and to correlate these findings to the knowledge structures developed by the participants. There are three primary research questions in Study 1:

1) Does the modeling pedagogy improve problem-solving performance?
2) Does the modeling pedagogy help participants to construct more "expert-like" knowledge structures as shown by card sort analysis?
3) How are the knowledge structures related in predicting problem-solving performance and Force Concept Inventory scores?
4) How are Force Concept Inventory scores and "expert-like" knowledge structures (as predicted by card sort analysis) related in predicting problem-solving performance?

3.2 Method

3.2.1 Participants

The study utilized first and second-year physics classes located in a Midwestern metropolitan area. Sixty-one first year modeling participants participated in the study. Twenty-seven of the modeling volunteers attended a suburban public school and thirty-four attended a suburban private school. A total of thirty-six non-modeling participants from two different schools volunteered for the first study. Twelve of the non-modeling volunteers attended a suburban private school while the other twenty-four attended a suburban parochial school. All of the first-year participants were enrolled in a trigonometry-based physics class. The two non-modeling classes were both traditional in the sense that they were more teacher-centered than a modeling class, used verification labs instead of exploratory labs and used a fair number of demonstrations which are not done to a large extent in the modeling classes. One of the differences between the two non-modeling instructors was that the private school instructor stressed misconceptions to a greater extent since the misconception research was well known to the instructor.

Two second-year advanced placement physics classes one from the public modeling school, totaling 22 participants, and one from the non-modeling private school, totaling 19 participants, volunteered for the study. Both advanced placement physics classes contained a mixture of students studying for the advanced placement calculus-based test (AP C) and the non-calculus based test (AP B). Both of the second-year classes focused on problem-solving in preparation for sitting for the AP exam in May. The major difference is that the modeling second-year's problem-solving focused on the identification of a model before deciding upon an algebraic or graphical solution method whereas the non-modeling group focused on algebraic solutions. All of the schools in the study had similar socio-economic backgrounds that support close to 100% of all graduates moving on to post-secondary training, primarily four-year degree programs.

3.2.2 Experimental design

Initially, the design was to involve all classes taking two paper and pencil tests and completing a card sorting task. The paper and pencil tests consisted of the Force Concept Inventory and a problem-solving task designed for this study. All classes were to take a pre and post test FCI, a pre and post-test card sort and a post test problem-solving task. Due to some initial difficulties with several of the schools, I was unable to complete all of the assessment tasks as initially planned. The tests completed and the number of participants in each group is summarized in Table 3-1.

TABLE 3-1 Summary of Assessment Tasks taken by each Experimental Group

Class	Pre – FCI	Post – FCI	Pre – Card sort	Post – Card Sort	Post Problem-Solving Task	Number of Participants
Private first year Modeling	Yes	Yes	Yes	Yes	Yes	34
Private first year Non-Modeling	No	Yes	No	Yes	Yes	12
Public first Year Modeling	No	Yes	No	Yes	No	27
Parochial First year Non-Modeling	No	Yes	No	Yes	Yes	24
Public second year Modeling	No	No	Yes	Yes	Yes	22
Private second year Non-Modeling	No	Yes	No	Yes	Yes	19

The pre-test card sorts that were completed could not be completed until between two weeks and one month into the school year depending on the class. Therefore, any results from pre to post-test must be analyzed accordingly.

3.2.3 Force Concept Inventory

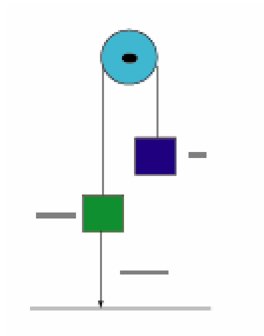
The Force Concept Inventory was administered to participants with the permission and guidelines utilized by Hestenes et al. (1992). All participants were given 30 minutes to complete the 30 question conceptual task covering topics in mechanics and kinematics. The post test was given to all groups within two weeks of completing all of the traditional kinematics, mechanics and energy concepts. The instruction scripts used for these tasks are included in Appendix A.

3.2.4 Card Sort Task Development

3.2.4.1 Pilot Card Sort Task

The card sort task was based on Chi et al. (1981). Chi et al. (1981) discovered that experts categorized problems based upon major physics principles such as conservation of energy and Newton's second law. Experts have a tendency to collapse problems into two fundamental principles (i.e., momentum and energy). While these principles are contained in the models students developed while using the high school modeling pedagogy there are several differences. For example, the momentum principle is divided into two models: impulsive force and constant force models. The six models utilized in this study were: constant velocity, constant acceleration, Newton's second law (constant force model), impulse and momentum, circular motion, and conservation of energy. Each of these models would also be developed in a non-modeling class but would be discussed in a very different manner. In the sorting task there were nine sets of problems designed with a series of six questions with isomorphic surface features but different deep structures required for solving. Each of the six questions required the deployment of a different model for the ultimate solution. A number of the sets were designed utilizing surface features such as pulleys, springs and inclined planes found by Chi et al. (1981) to be particularly enticing to novice students. For example, the following two problems were designed with similar surface features but require the deployment of different models for successful solution and actually use the same diagram:

> 1. Constant velocity model: A rope that passes over a massless frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. Due to a large amount of friction in the pulley the basket of bricks moves downwards at a velocity of 3 m/s. How far does the basket move in 5 seconds?



> 2. Momentum/Impulse Model: A rope that passes over a mass less frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. If the system, with a total mass of 25 kg, experiences a net force of 125 Newtons over a period of 3 seconds what is the change in velocity?

I decided to test the veracity of the Chi et al. (1981) finding that novice students have a knowledge structure based upon surface features by designing problems based not only upon incline planes, springs, pulleys, circles and free fall but also using more common surface features such as elevators, graphical features, sports, and transportation. The final surface features used in the pilot study were: inclined planes, springs, pulleys, free fall, circles, graphical features, baseball bats, elevators, and vehicles. The card sets were tested to determine how an "expert" modeler would sort the problems. Three experienced modeling instructors with experience ranging from 33 to 12 years were given the instructions and asked to sort the problems based upon solution path utilized. Two of the instructors sorted the scores as predicted based upon the models listed ignoring all surface features. One of the modeling instructors sorted one problem differently but upon questioning it was discovered that he had missed a key word that differentiated the models. The same set of cards was given to a teacher who had not been trained in modeling who sorted the cards in the same fashion as the modeling instructors. A pilot study was run in which the cards were given to two advanced placement second year classes (one modeling and one non-modeling) to test for differences. The students were told not to solve the problems but to sort them based upon similarity in solution path. The instruction script for the card sort is located in Appendix A (this script was used in the pilot study and Study 1). Based upon the results, several of the questions were changed and some additional questions added so the task so that its final version contained 58 problems. The final set of questions is located in its entirety in Appendix B.

3.2.4.2 Card Sort Scores

In order to achieve the purpose of this study I needed to correlate the FCI scores with the participants' card sort task performance. Therefore, I needed a way to quantify the sorting task performance in order to achieve an "expert-like" score. I initially set up a confusion matrix based upon expert model classification of the problems utilized. Each problem was assigned a number with the tens digit based upon the deep structure model used and tenths digit based upon the surface feature addressed by the problem. For example, problem 2.2 was given this value because its deep structure consisted of constant acceleration while it surface feature was based upon an inclined plane. Whereas, 3.2 would be considered a constant force model problem (i.e., Newton's second law) using a cover story of an inclined plane surface feature. Therefore, an expert modeler would sort all of the cards solved with the deployment of the constant velocity model together, and all of the constant acceleration model problems together, etc. I initially entered all the groupings completed in an expert matrix similar to the example in figure 3-1. An X is placed in the column under problem 1.1 next to each problem that a subject assigned to that category. For example, if a student at sorted problems 1.1, 1.2, 2.2 and 3.3 together there would have been an X placed in the boxes in the column under problem 1.1 at the location of 1.2, 2.2 and 3.3. An X would not have been placed in the location of 1.1 since one cannot sort problem 1.1 with itself. A more expert or deep structure sorting would have the majority of the groupings falling into the bordered boxes whereas a novice or surface feature grouping would fall outside of the bordered boxes as shown in figure 3-2 and 3-3.

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1									
1.2									
1.3									
2.1									
2.2									
2.3									
3.1									
3.2									
3.3									

Figure 3-1: Example of Expert Matrix used for card sort

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1		X	X						
1.2	X		X						
1.3	X	X							
2.1					X	X			
2.2				X		X			
2.3				X	X				
3.1								X	X
3.2							X		X
3.3							X	X	

Figure 3-2: Example of an Expert-like card sort

	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
1.1				X			X		
1.2					X			X	
1.3						X			X
2.1	X						X		
2.2		X						X	
2.3			X						X
3.1	X			X					
3.2		X			X				
3.3			X			X			

Figure 3-3: An example of a surface feature sort completed by a pure novice.

Of course, in the case of this study the majority of the matrixes did not show such clean cut examples as Figure 3-2 and 3-3.

3.2.4.2.1 Card Sort Expert Score

In order to test my predictions I needed a method by which I could obtain a numerical score for a students "expert-likeness" so that it could be correlated with their FCI scores. I determined that the best expert percentage score was obtained by dividing the number of hits in the expert boxes vs. the number of hits in and out of the expert boxes (then multiplied by one hundred). Therefore, a student with a card sort that was based upon models would obtain a value close of one hundred; whereas, a student sorting based strictly upon surface features would obtain a score as low as zero.

3.2.4.2.2 Card Sort Surface Feature Score

Based upon the pilot data I decided that it would be useful to complete an additional matrix based upon surface features in order to obtain a companion surface feature score. Figure 3-4 is an example of a surface feature matrix for a student completing a sorting task based completely on surface features. An X is placed in the column under problem 1.1 next to each problem that a subject assigned to that category.

	1.1	2.1	3.1	1.2	2.2	3.2	1.3	2.3	3.3
1.1		X	X						
2.1	X		X						
3.1	X	X							
1.2					X	X			
2.2				X		X			
3.2				X	X				
1.3								X	X
2.3							X		X
3.3							X	X	

Figure 3-4: An example of a surface feature matrix for a novice student sorting cards based upon surface features.

The surface feature percentage score was obtained by dividing the number of hits in the surface feature boxes by the total number of hits both in and out of the surface feature boxes (then multiplying by one hundred). It was determined based upon a qualitative analysis of the student card sorts that when sorting was based upon surface features students used physical items such as incline planes, pulleys, free fall, graphical features, circles and neglected transportation, sports and elevator features. Therefore, the final surface feature score utilized for comparison only included physical items used in Chi et al. (1981) and graphical features.

3.2.4.2.3 Card Sort "Questions-Asked" Score

During Study 1's analysis it was discovered that not only were students sorting by expert models and surface features but also based upon what the question was asking them to solve for hereafter referred to as the questions-asked strategy. This was somewhat unexpected as it was not mentioned specifically in past research. It was discovered after this experiment was conducted that Silver (1979, 1981) found that junior high school math students sorted problems via the questions-asked in each problem. Upon further review, the Chi et al. (1981) study did mention that novices do use literal features of the problem during card sorts. Is the questions-asked strategy what they are referring too when they said literal features? After this experiment was completed it was also discovered that Keith (1993) found the question-asked strategy was used by college physics students as reported in an unpublished doctoral dissertation. It is possible that the surface feature score is a very novice sorting and that as sophistication increases students might realize that problem solving based upon surface features does not lead to much success thereby developing a questions asked sorting. The questions-asked sorting could later be modified by students as their knowledge structures become more robustly expert-like with additional learning. It is possible that the only reason the questions-asked sorting was seen in this study was that the pre-card sort was given up to a month into the school year. Therefore, a third matrix was developed based upon questions-asked. The questions-asked score was determined for the questions-asked matrix using the same formula developed for expert and surface feature scores. Figure 3-5 shows an example of a questions-asked matrix. An X is placed in the column under problem 1.1 next to each problem that a subject placed in the same category. For example, in Figure 3.5 all of the problems in the first box (problems 1.1, 1.2, 1.9 and 2.5) might have all asked the students to determine the displacement of an object.

	1.1	1.2	1.9	2.5	4.9	1.7	2.1	5.3	6.8
1.1		X	X	X					
1.2	X		X	X					
1.9	X	X		X					
2.5	X	X	X						
4.9						X	X		
1.7					X		X		
2.1					X	X			
5.3									X
6.8								X	

Figure 3-5: Sample Questions-asked Matrix example demonstrating the non uniformity of the questions if grouped purely on the questions-asked

There were a total of seven different "types" of questions asked amongst all of the 58 problems. The percent of expert groupings that can be explained by the questions-asked groupings ranges from a high of 34% to a low of 5% depending on the question asked. For example, the greatest preponderance of questions asked the students to solve for the final velocity, initial velocity or average velocity. The questions-asked velocity grouping accounted for 34% of the expert model groupings. All of the other questions-asked groups accounted for approximately 17% or less of the expert model groupings which is very close to the percentage accounted for by a random grouping (i.e., 14%). Therefore, the data discovered from this matrix must be looked at carefully since the questions are biased in certain directions since the problems were only designed to be isomorphic around models and surface features. However, the results could lead to interesting comparisons between groups.

3.2.5 Problem-Solving Task Development

The problem-solving task was designed after the pilot card sort in order to determine if an increased expert-like card sort score correlated with increased performance not only on qualitative measures such as the FCI but also on a quantitative problem-solving task (otherwise known hereafter as the PS Task). The PS Task consisted of seven quantitative problem-solving tasks. There was one question each for the six models using in this study: constant velocity, constant acceleration, circular motion, energy, Newton's second and third law, and impulsive force. The seventh task was a graphical constant velocity/constant acceleration task. Three of the tasks (circular motion, Newton's second law, and energy) were adapted from the Mechanics Baseline test with permission from David Hestenes. The problems were designed so that all of the major models covered in the Modeling curriculum were covered and based upon problems that all physics textbooks traditionally cover. The graphical task was added so that it could be determined if participants in both conditions could solve problems utilizing graphical methods in addition to algebraic methods. The PS Task was given within two weeks of completing all of the traditional kinematics, mechanics and energy concepts in the course currently taken by the students. The PS Task in its entirety is located in Appendix C.

3.2.6 Analysis of data

Based on the data above, I was able to correlate the FCI scores with the PS Task scores to determine if my participants were in line with the data from previous studies that showed that FCI scores correlated well with problem-solving skills as determined by the MBT. In addition, I was able to connect knowledge structure and expert/novice research to modeling pedagogy by correlating the expert, surface feature and questions-asked scores with the FCI and the PS Task.

3.3 Hypotheses and Predictions:

Hestenes et al. (1992) and Hestenes and Wells (1992) determined that modeling participants outperform non-modeling participants on the Force Concept Inventory (FCI) and Mechanics Baseline Test (MBT). Study 1 is designed to extend these finding by discovering if there is a connection between the enhanced performances on conceptual and problem solving based tests and the knowledge structures developed by the participants in two conditions: modeling and non-modeling classes. One set of participants were taught physics with non-modeling approaches while the other set of participants were taught with the modeling pedagogy as described in Chapter 1. There has not been any past research into what cognitive resources the modeling students process that allow for this performance. Study 1 will expand on past research by determining the differences in knowledge structures between the two groups and any correlations that may exist between students' knowledge structures and FCI performance by having the participants complete a card sort task similar to the one conducted by Chi et al (1981). Based on my understanding of the literature I make the following predictions:

- I predict that the first and second year modeling participants would out-perform the non-modeling participants on the FCI and a problem-solving task based upon the MBT.
- I expect to see a gain in FCI scores from pre to post test for the modeling classes.
- I predict that due to the modeling pedagogy the card sort groupings of the modeling participants will be based more on deep structure (i.e., models) than the non-modeling students whereas the non-modeling students will base their card sort groups more on surface features.
- For the modeling courses I expect to see a gain in the use of deep structure and a lessening of the dependence on surface features from pre to post test.
- I predict that there will be a correlation between knowledge structures observed and test performance, such that, a more expert-like knowledge structure will allow for better performance on the paper and pencil tests utilized.
- I predict that second-year students will demonstrate significantly better performance on all tests regardless of course type from that of first-year physics students since they have experienced an additional year of study in physics concepts.

3.4 Results

3.4.1 Results for First year Trigonometry based Physics classes

3.4.1.1 Comparison of the two Non-Modeling Classes and two Modeling Classes

The scores on all post tests and the significance within groups are summarized in Table 3- 2. Using a one-way analysis of variance (ANOVA) I found significant differences between the non-modeling participants' post-test FCI scores and the participants post-test surface feature scores for the two non-modeling classes. The post-test FCI scores for the private school were surprising since they did not fall in line with previous research such as Hestenes et al. (1992) who found traditional non-modeling classes scored 50% or lower on the FCI.

In addition, Hestenes and Wells (1992) found that a high FCI is necessary but not sufficient for a good score on a problem-solving task such as the MBT. Therefore, normally one would expect that a higher FCI score would allow for a higher PS Task score. However, this data demonstrates that the PS Task Scores between the two non-modeling classes are not statistically significant even through the post-FCI scores were.

A one-way analysis of variance (ANOVA) revealed no significant differences between the two modeling classes. The one score that could not be compared was the PS Task score since the public modeling class could not be given that assessment.

Table 3-2 Summary of First Year Classes Average Post-Test Percentage Scores and the Significance Levels within Pedagogies. Scores that are statistically significant are highlighted in pink. NS stands for non significant.

Class	FCI	FCI F-statistic and p-value	Card Sort – Expert Score	Expert F-statistic and p-value	Card Sort – Surface Feature Score	Surface Feature F-statistic and p-value	Card Sort– Questions -asked Score	Questions Asked F-statistic and p-value	PS Task Score	PS Task F-statistic and p-value
Non-Modeling										
Parochial	39	$F(1, 34) = 27$, $p < 0.001$	30	NS	51	$F(1, 34) = 9.08$, $p < 0.005$	23	NS	42	NS
Private	68		33		30		33		50	
Modeling										
Public	65	NS	41	NS	21	NS	40	NS		
Private	65		44		18		34		62	

3.4.1.2 Comparison between all of the First-year Physics Classes

For further data analysis, the post-test scores that were not statistically significant between the two non-modeling groups (expert card sort, questions-asked, and PS task scores) were collapsed. The F-statistics and p-values found from utilizing a one-way ANOVA to determine statistical significance of these three scores between pedagogies on collapsed scores are summarized in Table 3-3. The F-statistics and p-values found from utilizing a one-way ANOVA to determine statistical significance for surface feature and FCI post-test scores between the collapsed modeling group and uncollapsed non-modeling classes are also summarized in Table 3-3.

Table 3-3 Summary of ANOVA results between Modeling and Non-Modeling Class Post-Test Scores. Scores that are statistically significant are highlighted in pink. NS stands for non significant.

Comparison	F-statistic	p-value
Expert card sort score for collapsed non-modeling class vs. collapsed modeling classes	F (1,95) = 18.74	p < 0.001
Surface feature card sort score for Parochial Non-Modeling class vs. collapsed modeling classes	F (1, 83) = 82.17	p < 0.001
Surface feature card sort score for Private Non-Modeling class vs. collapsed modeling classes	F (1, 71) = 9.2	p < 0.003
Questions-asked card sort score for collapsed non-modeling class vs. collapsed modeling classes	F (1, 95) = 3.11	p < 0.081
FCI score for Parochial Non-Modeling class vs. collapsed modeling classes	F (1, 83) = 31.81	p < 0.001
FCI score for Private Non-Modeling class vs. collapsed modeling classes	F (1, 71) = 0.32	NS
PS Task Score for collapsed non-modeling class vs. collapsed modeling classes	F (1,67) = 10.76	p < 0.002

These results show the expert card sort scores between pedagogies are significant. The modeling group has a higher expert-like knowledge structure. A qualitative analysis of the card sort shows that a number of students (especially modeling students) were collapsing a number of the models by the end of the school year. For example, a number of students collapsed constant velocity model with constant acceleration model which is not surprising since constant velocity is a special case of zero constant acceleration. This type of collapse is actually demonstrating a more expert like view since all of the models could be collapsed into two fundamental principles; i.e., energy and momentum. However, these types of collapses actually tend to depress the expert score as it is presently being calculated.

Even though the two non-modeling classes differ on their surface feature dependence by the end of the year it was found using a one-way ANOVA that both non-modeling surface feature post-test scores differed significantly from that of the modeling group. This finding demonstrates that the non-modeling group as a whole is more surface feature oriented than the modeling group.

The questions-asked scores across all groups were marginally significant with the modeling group demonstrating a larger score by the end of the year. No predictions were made about the questions-asked score since at the time the study was conducted it had not been discovered by the author that this type of card sort had been a finding in prior research studies. A larger score on the part of the modeling classes might be caused by the decline in surface feature score and the attempt on the part of the students to use a more sophisticated strategy while on the way to becoming more "expert-like" as discussed in Section 3.2.4.2.3.

The surprising finding was that the FCI post-test score for the private non-modeling class was not significantly different from the modeling FCI post-test score while the FCI post-test score for the parochial non-modeling class was significantly different from the modeling FCI post-test score. This may have been due to the private school non-modeling instructor's stress on misconceptions.

Due to past research one might expect the PS task score for the private non-modeling group to be more comparable to the modeling groups' score given the high FCI score but this is not the case. The PS task scores for the all the non-modeling groups were significantly lower than the modeling groups. In addition, the PS Task results demonstrated that none of the participants (modeling or non-modeling) differed in their ability to solve problems graphically based upon the two graphical problems contained on the test. As shown by the Chi-square 2 X 2 table in Table 3-4 the difference in error rate on these graphical problems was not significant.

TABLE 3-4 Solvers vs. Non-Solvers of Graphical Problems contained on the PS Task

	Modelers	Non-Modelers	Total
Correct Solutions	64	54	118
Incorrect Solutions	16	9	25
Total	70	73	143

(Chi-square test = 2.75, $p < 0.1$, df = 1)

3.4.2 Results for Second-year Advanced Placement Physics Classes

The one post-test score that could not be compared between the two second-year physics groups was the FCI post-test score as the FCI score for the public modeling class was unattainable. The scores for all other post-tests taken by both second-year physics classes were compared using a one-way ANOVA to determine statistical significance. The private and public second-year advanced placement (AP) modeling class scores on all post tests and the F-statistics and p-values obtained is summarized in Table 3-5.

Test	Non-Modeling Second-Year Post-test Score	Modeling Second-Year Post-test Score	F-statistic	p-value
FCI Score	67			
Card sort – Expert	32	51	$F(1,39) = 22.57$	$p < 0.001$
Card Sort – Surface feature	29	18	$F(1,39) = 9.16$	$p < 0.004$
Card Sort – Questions-asked	24	30	$F(1, 39) = 7.56$	$p < 0.009$
PS Task Score	48	76	$F(1,38) = 23.04$	$p < 0.001$

Table 3-5 Summaries of Post-Test Percentage Scores and ANOVA comparisons between Pedagogies for Second-Year AP Modeling and Non-Modeling Class. Problems that are statistically significant are highlighted in pink.

All of the findings are in line with the original predictions that the Modeling second-year AP class would outperform the non-modeling second-year AP class. I did not have any predictions initially for the questions-asked score as this was a surprise finding. The non-modeling AP class had a lower questions-asked score than the modeling AP class which is not what one might expect if one considers questions-asked to be a type of surface feature. However, this trend is similar to that seen with the first-year course score comparisons in Section 3.4.1.2. Therefore, the questions-asked sorting might represent a shift from surface feature towards more problem-solving sophistication.

3.4.3 Comparison of First-year and Second-year Physics Classes by School

A comparison was made between the post-test scores for the first year trigonometry based class and its companion AP second year class. One must not set great value on these comparisons because they are not looking at the scores made by the same students. However, this case study is able to compare average post-test gains between the first year and second year classes at the private non-modeling school and the public modeling school. In addition, the first and second-year classes at each school were taught by the same instructors. The second downside to this comparison is that the modeling classes were only able to be compared on their card sort scores because of unavailability of a second-year modeling FCI score and a first-year modeling PS Task score. Table 3-6 summarizes the average post-test scores for the non-modeling school's first and second year classes as well as these same values for the modeling school.

Table 3-6 Summary of First-year vs. Second-year Physics Classes Average Post-Test Scores and Statistics by Pedagogy. Problems that are statistically significant are highlighted in pink. NS stands for non significant.

	Non-Modeling Groups				Modeling Classes			
Test Scores	**First –year**	**Second – year**	**F-Statistic**	**p-value**	**First –year**	**Second – year**	**F-Statistic**	**p-value**
FCI Post-test Score	**68**	**67**	F (1,30) = 0.205	NS	**65**			
Post-test Card sort – Expert	**33**	**32**	F (1,30) = 0.062	NS	**41**	**51**	F (1,48) = 5.78	p < 0.02
Post-test Card Sort – Surface feature	**30**	**29**	F (1,30) = 0.027	NS	**21**	**18**	F (1,48) = 0.667	NS
Post-test Card Sort – Questions-asked	**33**	**24**	F (1,30) = 0.218	NS	**40**	**30**	F (1,48) = 3.177	p < 0.081
PS Task Score	**50**	**48**	F (1,30) = 0.194	NS		**76**		

Using a one-way ANOVA it was found that the post-test scores between the first year and second year non-modeling classes are not significant as shown in Table 3-6. The second year non-modelers showed no significant improvement in FCI, PS Task or any post-test card sort scores. Indeed, the only score in which the averages point towards the original prediction of an improvement in scores was the questions-asked score but it is not even marginally significant. The PS Task results are quite surprising given the class focus on AP problem-solving test.

A second comparison using a one-way ANOVA completed for the modeling first and second year classes showed that only the expert card sort post-test score was statistically significant while the questions-asked score was marginally significant. The second year modelers showed no significant changes in the surface feature score from the first-year modeling group. This shift in post-test questions-asked scores from first-year to second-year proved to be only marginally significant but lower for the second-year class. At the same time that the questions-asked score declines the second-year expert score increases which might be pointing towards a possible shift in sophistication from surface feature to questions-asked to expert sort.

3.4.4 Comparison of Pre-test to Post-test results

3.4.4.1 Pre-test to Post-test results for Private school First-year Modeling group

Pre and post-test results for the FCI and all card sort matrixes were obtained from the private school modeling group. The results shown here probably only show a portion of the actual gain that must have occurred since the pre-test card sort task was not given to the class till a month into the school year due to administration difficulties. A one-way ANOVA was utilized to determine if the differences in the pre and post test averages were significant. Table 3-7 shows a summary of the pre to post-test averages for the class and the p-values and F-statistics discovered.

Table 3-7 Summary of Pre to Post-test Percentage Scores and ANOVA results for the First-year Modeling Private School Class. Problems that are statistically significant are highlighted in pink. NS stands for non significant.

Test	Pre-test Score	Post-test Score	F-statistic	p-value
FCI Score	27	65	$F(1,64) = 91.45$	$p < 0.001$
Card sort – Expert	28	44	$F(1,64) = 38.03$	$p < 0.001$
Card Sort – Surface feature	19	18	$F(1,64) = 0.45$	NS
Card Sort – Questions-asked	46	34	$F(1, 64) = 6.64$	$p < 0.012$

The results are in line with the predictions in all areas except for the surface feature scores. The modeling students' results show that the students gained significantly in conceptual understanding as shown by the FCI and in expert knowledge organization as shown by the cart sort expert score. One must remember that the pre-test was given one month into the school year which means that the differences would probably have been greater if the pretest had been administered earlier. In addition, at the time of the pre-test the class had just started work on the constant acceleration model and the qualitative results of the pre-test sort show that a number of students sorted the constant velocity and some constant acceleration cards together regardless of surface features but sorted the rest of the cards based upon surface features or questions-asked. It is possible that during development of the first model the modeling students quickly discovered that surface features should not be relied upon and switched to a questions-asked categorization hence the reason why this sorting was observed in this study. Therefore, the combination of modeling pedagogy and late pre-test could have resulted in a ceiling effect in terms of surface features thus producing a decreased range effect.

3.4.4.2 Pre to Post-test Results for Public School Second-year Modeling Class

I was able to obtain pre test card sort data for the second-year public school modeling group. Therefore, comparisons can be made between this data. A one-way ANOVA was utilized to determine if the differences in the pre and post test averages were significant. Table 3-8 summarizes the pre to post-test averages for the second-year modeling class the p-values and F-statistics discovered.

Table 3-8 Summary of Pre to Post-test Percentage Scores and ANOVA results for the Second-year Modeling Public School Class. Items that are statistically significant are highlighted in pink. NS stands for non significant.

Type of card sort	Pre-test Score	Post-test Score	F-statistic	p-value
Card sort – Expert	40	51	F (1,41) = 5.82	p < 0.02
Card Sort – Surface feature	34	18	F (1,41) = 8.55	p < 0.006
Card Sort – Questions-asked	33	30	F (1, 41) = 0.54	NS

This information falls in line with the initial predictions with the exception of the questions-asked score for which there were no prediction. Even through the questions-asked score is not significant it is showing a decline from pre to post test like the first-year modeling group scores while showing an increase in expert score above that seen in the first-year scores (21.7% relative difference)*. In addition, due to uncontrolled difficulties the pre-test card sort task was not given till two weeks into the school year which may have caused the pre-test expert scores to be elevated in value while producing a deflation in the surface feature and questions-asked score but not to the same degree as the first-year group.

3.4.5 Post-test Score Correlations

The data from all of the groups were collapsed to determine if the initial predictions concerning whether more expert-like knowledge structures allow for better performance on the FCI and the PS Task were correct. I performed correlation studies on the data from all groups and then the same correlations excluding the data from the private school non-modeling first year group. This was done because of the discrepancy from past research that was shown between this non-modeling group's FCI and PS Task scores. Any FCI correlations do not include the modeling second year class since FCI data was not obtained for it. In addition, for any correlations with the PS task score the first year public school modeling group has been excluded as that data was not obtained. In past research it was shown that the FCI was highly correlated with problem-solving abilities on the MBT test (Hestenes et al., 1992; Hake, 1998; Desbien, 2002). To determine if my data is in line with the prior research I determined the correlation between FCI post-test scores and PS Task Scores. The correlation coefficient between the FCI and the PS Task is 0.65 if the data from the private non-modeling group is excluded due to its inconsistencies. If no data is excluded the correlation coefficient between the FCI and the PS task is 0.62. These two correlations are not significantly different as determined by a Fisher r-to-z transformation ($z = 0.26$, $p < 0.795$).

---

*Relative difference was calculated by subtracting the lowest result from the highest result, dividing by the average of the two results and then multiplying by 100.

However, in either case the data is in line with the correlation coefficient value of 0.68 obtained by Hestenes and Wells (1992) showing that the FCI score accounts for 43 to 38% of the variance in PS Task score, which is very similar to the 46% of variance discovered by Hestenes and Wells (1992). Having determined that the PS Task and FCI scores are in line with past research I performed correlation tests with all of the data obtained in this study to determine if any of the card sort scores were predictive of the FCI and/or PS Task scores. Table 3-9 summarizes the results of these analyses.

As predicted the correlations are quite high for expert score vs. FCI score and expert score vs. PS Task score. The correlations are even higher between expert score and the FCI score when the non-modeling private school data is excluded, 0.64 vs. 0.57 respectively. These values are not significantly different as determined by a Fisher r to z transformation. The correlations between expert and the PS Task score are almost identical with and without the private school non-modeling data. The data above does imply that the methods used in the private non-modeling school seem to inflate the FCI scores while not increasing the participants "expert-likeness" as measured by the card sort scores. However, the scores of the non-modeling group are as highly correlated with the PS Task as the other schools and very predictive of final PS Task score. Surface feature or questions-asked scores have a negative correlation for both FCI and PS Task scores but only surface feature is significant. Therefore, the surface feature score is more predictive of the variance observed in both post-tests. The surface feature score accounts for about 25% of the variance seen in the FCI scores and 22% of the variances observed in the PS Task scores but in a negative direction. Therefore, the higher the surface feature score the lower the FCI and PS Task scores.

Correlations were also completed between the different card sort scores. It was found that the surface feature score and the expert score was negatively correlated with a correlation value of 0.50 which accounts for 25 % of the variance in the expert score. All of the correlation values between the card sorts were the same whether the non-modeling private school was included or not. As one can see most of the post-test scores are correlated with each other. However, the questions-asked score is the exception being the only post-test which is not significantly correlated with any of the other scores accounting for less that 7% of the variance in all cases. The correlation values for all data collected are summarized in Table 3-9.

Because of the high correlations between the FCI, PS Task and expert scores a stepwise regression was completed to compare the combined expert card-sort and FCI scores correlation with the PS Task score. It was found that together the expert score and FCI score accounted for 44% of the variance in PS Task score. This combination allows one to account for an increase in the variance of only 6%. Therefore, it seems that the FCI and the expert score are strongly correlated to each other and are accounting for a similar effect on the PS Task score. Therefore, a high expert score leads to a high FCI score but not necessarily the other way around while both of these lead to a high PS Task score, as supported by the high correlation between the two scores.

Table 3-9 Summary of Correlation Coefficient/r^2 values between all post-tests. Items that are statistically significant are highlighted in pink.

Test	Correlation Coefficient (R)	R^2 (%)
FCI vs. PS Task	0.62	38
Expert score vs. FCI	0.57	33
Expert score vs. PS Task	0.65	41
Surface feature score vs. FCI	- 0.50	25
Surface feature score vs. PS Task	- 0.45	20
Surface feature score vs. Expert Score	- 0.50	25
Questions asked score vs. FCI	- 0.24	6
Questions asked score vs. PS Task	- 0.13	2
Questions asked score vs. Expert Score	0.08	0.06
Questions asked score vs. Surface Feature Score	0.24	6

3.5 Discussion

Study 1 was designed to determine the effect modeling pedagogy had on the knowledge structures constructed by the participants and their correlation to FCI and PS task scores. In addition, study 1 was designed to determine if students producing higher FCI scores had more "expert-like" knowledge structures than those producing lower FCI scores and if these scores were correlated highly with the PS Task score. This study not only allowed for the comparison between all the participants' scores but also allowed for a comparison of the post-test abilities between participants trained using the modeling pedagogy vs. non-modeling pedagogies at both the first-year and second-year levels.

3.5.1 Research Question 1

Study one was designed to answer the following question:

> Does the modeling pedagogy improve problem-solving performance as predicted by the prior research into the efficacy of the pedagogy?

A comparison of the FCI and the PS Task scores between groups demonstrates the success of the modeling pedagogy as shown in prior research (Hestenes et al., 1992; Hestenes and Wells, 1992; Hake, 1998; Desbien, 2002; Vesenka et al., 2002; and Brewe, 2002). The only anomalous finding in study 1 occurred with the FCI scores between the modeling groups and the private non-modeling group. The FCI data was not totally in line with past research since the private non-modeling group's FCI scores were not significantly different from the modeling groups while the parochial non-modeling group's FCI scores were highly different (a 50% relative difference). However, the data for all groups are in line with past research in the area of the PS Task score demonstrating that modeling participants produced significantly higher scores on this problem-solving task (a 29% relative difference). The private non-modeling teacher was very well versed in the research on misconceptions and by having the participants repeatedly look at misconception examples this could have allowed for this group's production of higher than usual FCI scores while not producing better PS task scores. Research data from Arizona State University has shown that highlighting misconceptions in this fashion can cause an increase in the FCI scores over the short term (private communication with David Hestenes). In the case of the private non-modeling class the higher FCI scores did not enhance their problem-solving ability as was the case with the other three classes.

An increase in conceptual understanding on the part of the modeling students can be observed dramatically by looking at the increase in their FCI score from pre to post test. The first-year modeling students demonstrated an increase in FCI score from 27% to 65% by the end of the school year. This increase represents an 83% relative difference pre to post test.

The strength of the modeling pedagogy was again proven by an analysis of the differences in the post-test scores between second-year modeling and non-modeling groups. The second-year modeling group had an average score of 76% on the PS Task after two years of the modeling pedagogy while the non-modeling second year group scored an average of 48% - a 45% relative difference. Since both second year courses were AP courses they were both covering the same syllabus and stressing the problem-solving needed to score well on the AP exam at the end of the school year. However, it would seem that the identification of a model as prescribed in the model pedagogy is producing significantly better performance.

When comparing the post PS Task between the first-year and second-year non-modeling groups one finds that there was not a significant increase their performance. It would almost seem that the second year of non-modeling physics training does not help the students to improve. However, when comparing the modeling groups' scores from first to second year one must take care since the first year group was unable to complete the PS Task and the second year group could not complete a post-FCI test. The first year modelers as an entire group has a post-test FCI score of 65% which was not significantly different from the private school non-modeling group while the second-year modeling group produced a 76% average PS Task score... Given that the modeling participants FCI and PS Task scores are highly correlated the FCI score one would predict for the second year modeling group would be well over 83; thereby, demonstrating a significant increase between the first year modeling post-test FCI and the second year modeling post-test FCI. Of course, one must take care when making this type of comparison since these were two different sets of students and the data is not part of a longitudinal study.

These results point to the fact that the modeling students showed enhanced performance above that of the non-modeling groups on the FCI and PS Task in all comparisons except for the FCI score for the non-modeling private school course. Therefore, the prediction that the modeling students would demonstrate enhanced problem-solving performance was supported.

3.5.2 Research Question 2

While it is good that the data supports the past research and demonstrates that the two different populations are behaving as one would expect due to the pedagogies utilized in the classroom it does not begin to explain why this might be the case. Study 1 was also designed to answer the following question:

> Does the modeling pedagogy help participants to construct more "expert-like" knowledge structures as shown by card sort analysis?

Study 1 demonstrated that the modeling groups produced participants who had significantly higher expert card sort scores. This means that after the first year course the modeling students have a more expert-like knowledge structure then the non-modeling groups demonstrating a 29.7% relative difference between the two groups. One would expect a first-year participant who is more "expert-like" to rely less heavily on surface features and the data bears this out demonstrating that the modelers have significantly lower surface feature card sort scores than the surface feature scores on average for the non-modelers – a relative difference of 70%. This trend was also seen in the second-year group with the modelers producing an expert score that was higher on average than the non-modelers (45.8% relative difference) and a surface feature score that was lower on average than the non-modelers (a 46.8% relative difference). All of the results above were significant.

The surprise finding in the card sort analysis was that the students rely quite heavily on the questions-asked in the problem when attempting to sort them. This questions-asked score for the first-year students was the only card sort score that was only marginally significant between the two main populations; but, the modelers always produced the higher value. This was a surprise as it had not been specifically mentioned in the Chi et al. (1981) study. However, in that study they do define surface features to also mean "the literal physics terms mentioned in the problem". It is possible that this might be akin to the questions-asked strategy used by a number of students in this study. It was discovered subsequently that a questions-asked strategy was used by junior high math students and college level physics students (Keith, 1993; Silver, 1979, 1983). One could postulate that the three card sort scores are linked on a continuum such that a novice student might progress from a high surface feature score to a high questions-asked score as this is a more sophisticated strategy once they notice that a reliance on surface features does not produce high performance. Then as the novice student starts to develop a more "expert-like" knowledge structure the questions-asked score declines while the expert score increases. There seems to be qualitative evidence to suggest that this might be the case since a number of the first year modelers on the pre-card sort were grouping the constant velocity model problems together regardless of surface features but grouping other problems based on questions-ask or surface features. Of course, one must be careful with the questions-asked results since the cards were not designed to test that hypothesis therefore the types of questions-asked were not evenly divided across all surface features and all models.

It is also important to determine if the modeling students' knowledge structure shows a shift form surface feature leanings to a more expert-like knowledge structure. The determination of this was problematic due to the lateness of the pre-card sort meaning that it could have shown a decline in surface feature. The actual surface feature score is not significant for the first-year modelers' pre to post test. However, since the test was given a month into the school year it is possible that the modeling pedagogy caused a swift discarding of this organization due to being immediately shown that it was an unproductive organization. This is what might have led to the large pre test questions-asked score. The questions – asked score did show a significant decrease pre to post test (a 30% relative difference) while the expert score showed a significant increase pre to post test (a 44.4% relative difference). The expert score shift definitely demonstrates a shift in the reliance and knowledge structure towards deep structure.

A similar trend in the card sort scores is observed in the pre to post test scores for the second-year modeling group. The expert score pre to post test demonstrates a significant increase in the students' expert-like knowledge organization such that the students showed a relative increase in expert-likeness of 24%. The second year group also showed a significant decline in surface feature reliance – a 61.5% relative difference pre to post score. These scores seem to be demonstrating that knowledge organization of the second-year physics students showed a shift from a knowledge organization based upon surface features to one based upon models or deep structure. In addition, the second –year group showed a marginally significant decline in questions-asked score – a 9.5% relative difference. Since the pre card sort task was given much earlier into the school year for the second-year group these scores support the prediction that the knowledge organization might shift from the beginning of the year from a surface feature organization to one based upon questions-asked till the students reach a more expert-like organization at the end of the year. This would also give added support to the prediction that the reason the first-year group showed such a large questions-asked score might be due to the late administering of the test and happened to catch the transition from surface feature to questions-asked.

One could ask: Does the data show an improvement in the "expert-likeness" of the participants as they progress from first to second year? This is true for the modeling participants but not the non-modelers. The non-modelers did not become more expert-like from first to second year. Nor did they show a decline in reliance on surface features during problem solving as there was no significant difference in the post-test scores from the first to the second year. However, the modelers showed a significant increase in expert score during the second year with a 21.7% relative increase between the first-year modelers' and second-year modelers' post-test expert card sort. While there is no significant change in the surface feature and only a marginally significant change in the questions-asked scores from the first to the second year the decline in the questions-asked scores is almost a 10 percentage point decline (a 28.6% relative difference). This decline follows along with the hypothesis that the questions-asked score is a link between a surface feature based knowledge structure and the more "expert-like" knowledge structure since there is a similar difference observed in the expert score.

These findings definitely support the original prediction that the modeling students would should a gain in deep structure and a lessening in surface features pre to post test. In addition, the findings also support the conclusion that the modelers' knowledge structures are based more on deep structure while the non-modelers' knowledge structures are based upon surface features.

3.5.3 Research Question 3

Of course, one would hope that as one's knowledge structure becomes more "expert-like" than one would become better at problem solving. In order to determine if that might be the case, Study 1 was designed to answer the following question:

> How are the knowledge structures related in predicting problem-solving performance and Force Concept Inventory scores?

The design allowed for correlations between card sort scores, FCI scores and PS Task scores to be determined. The correlations showed that the higher the expert score the higher the FCI and PS Task score, accounting for 33% and 41% of the variance, respectively. The correlations between the FCI scores and the PS Task scores for all participants showed that the score obtained on the FCI accounts for a large percentage of the variance in PS Task score (i.e., in order to obtain a high PS Task score one needed a high FCI score). The data demonstrated that the PS Task created for Study 1 yields similar results as those obtained by Hestenes and Wells (1992) with the MBT. A high surface feature score was negatively correlated with both post-tests but accounted for a greater amount of the variance on the FCI test (25% vs. 20%). The expert and surface feature scores are negatively correlated accounting for 25% f the variance observed in each and this is in the direction one would expect. As the student becomes more expert-like they should rely less on surface features and more on deep structures. The questions-asked score did not seem to have much predictive value for either post-test and was not significant. In addition, the questions-asked score was not highly correlated to either the surface feature score or the expert score.

This analysis seems to imply that there are methods by which a teacher can achieve a large FCI score but not affect the "expert-likeness" of the students. When the stress on raising the FCI scores is removed from the classroom the "expert-likeness" of the students becomes highly predictive of their final scores on both the FCI and the PS Task while the FCI score then becomes highly predictive of the PS Task score.

The results definitely demonstrate that as predicted the type of knowledge structure observed correlates highly with the conceptual and problem-solving performance of the students. If the students rely on surface features it will depress their performance scores. However, if the students have a knowledge organization based upon deep structures it enhances their performance scores in both the areas of conceptual understanding as well as problem-solving performance. The questions-asked sort does not seem to correlate highly to the performance scores. This might be caused for two reasons. One reason is that the card sort problems were not isomorphic across questions asked. The second reason is that the questions-asked score simply does not correlate well. A surface feature knowledge structure seems to produce poor problem solving performance. Therefore, students might find it worthwhile to reorganize their structure. If they chose to reorganize their knowledge structure based upon questions-asked it might not help or hinder the performance to the extent that surface features and deep structure do so.

3.5.4 Research Question 4

The final research question in Study 1 was:

> How are Force Concept Inventory scores and "expert-like" knowledge structures (as predicted by card sort analysis) related in predicting problem-solving performance?

In order to answer this question a stepwise regression completed between the FCI and expert scores vs. the PS Task showed an increase in the variance was not accounted for when the scores were combined. The FCI and expert scores seem to be predicting a similar factor that is accounting for the same variance in the PS Task score.

3.5.5 In the Final Analysis

The previous discussion clearly demonstrates that the modeling pedagogy allows for the production of a more "expert-like" knowledge structure to develop which allows the participants to perform better on both qualitative and quantitative tests. Study 1 is only the first step to explaining how the modeling pedagogy is more successful than traditional methods. This study does not show how the more "expert-like" knowledge structure enables the participants to accomplish this increase in performance. In addition, this study does not allow one to determine what cognitive and metacognitive abilities the pedagogy fosters in the participants so that they demonstrate increased problem-solving ability.

Chapter 4
Study 2: Verbal Protocol Analysis of Problem-Solving and Metacognitive Behaviors

4.1 Purpose:

Study 2 was designed to compare the problem-solving strategies of the modeling and non-modeling participants in order to determine differences that lead to increased problem-solving abilities. The research questions for Study 2 are:

1) Will the "expert-like" knowledge structures of the modeling participants discovered in Study 1 allow for more "expert-like" problem-solving strategies as described in Table 2-1?
2) What metacognitive differences exist between the two groups due to the modeling pedagogy?
3) Will the modeling participants commit fewer physics errors while catching a larger proportion of those errors?
4) How do the results of Study 2 correlate with the findings from Study 1?

4.2 Method

4.2.1 Participants

Study 2 utilized first year high school physics classes located in a metropolitan area. Nineteen first year modeling students participated in the study. Four of the modeling volunteers attended a suburban public high school while the remainder attended a suburban private high school. A total of eleven non-modeling participants from a suburban parochial high school volunteered. All of the first year participants were enrolled in a trigonometry based physics class. All of the schools in the study had similar socio-economic backgrounds that support close to 100% of all graduates moving on to post-secondary training, primarily four-year degree programs.

4.2.2 Experimental design

Study 2 utilized a talk-aloud protocol. All of the participants were required to solve a series of five physics problems while talking aloud. Each subject was video taped. The video tape was situated so that it collected not only voice recordings but a record of the solution path recorded by the student. The students recorded their solutions on large sections of whiteboard using dry erase markers. The participants were allowed to refer to textbooks and notebooks. The participants were given unlimited time but if they seemed to reach an impasse, it was suggested that they move onto the next problem. After the participants completed the series of five problems a retrospective interview was conducted. The participants were shown the whiteboard they used for each problem and were asked to discuss all that they could remember thinking when solving that problem. Afterwards, clarification questions were asked of each subject before moving onto the next problem. A verbal protocol instruction script was used to ensure that all participants heard the same directions. The script can be found in Appendix D. The verbal protocol data was then transcribed and coded using a coding scheme (found in Appendix F) developed for this study.

4.2.3 Verbal Protocol Problem Development

The series of five physics problems was designed so that each problem required the deployment of a particular physics model. Four of the problems were designed to test the following models: constant acceleration, constant force, impulsive force and conservation of momentum, and energy. There was an additional problem that required the students' to deploy two models during the solution of the problem (constant acceleration and constant velocity models) and was graphical in nature. This particular problem was actually three separate problems in one. Therefore, technically the participants solved seven problems not five. The problems contained confounding information such as unnecessary information, information that suggested alternative models, or graphical information. For example, the following problem was the third verbal protocol problem out of the sequence of five.

> A comet has a mass of 50,000 kg and is in an elliptical orbit whose long axis is 6×10^{12} meters, and its period of revolution around our sun is 100 years. An asteroid has a mass of 10,000 kg and is in a circular orbit whose diameter is 2.2×10^{12} meters, and its period of revolution around our sun is 20 years. The comet has an x component of velocity of -5,000 m/s. It collides head on with the asteroid; whose x component of velocity is +11,000 m/s. After the collision the comet has an x component of velocity of +3,000 m/s. What is the final x component of the velocity of the asteroid?

This problem was designed such that it could activate either the circular motion model or gravitational field model for students while requiring the deployment of the impulsive force model.

The initial problems were given to four pilot participants to test for their solvability and the participant's ability to produce useful protocol data. Based upon the pilot test minor revisions were made to these problems. An additional two problems were added to the protocol in order to include the energy model and solution flexibility problem. The solution flexibility problem was designed so that it was more easily solved via a graphical approach in order to test the students' flexibility with utilizing alternate solution representations. The five problems can be found in their entirety in Appendix E.

4.2.4 Analysis of data

4.2.4.1 Segmenting Data

The transcriptions of the audio tapes for each participant were correlated to the video in order to produce a record of the physics path taken by each subject in each problem. The audio transcription was segmented into chunks corresponding to transitions in the physics problem-solving path. The following transcription segment was taken from a non-modeling subject (information in brackets contains notes made by the transcriber):

> Okay so I'm thinking is uh… v_F equals v_I, one of those type of problems, um, UAM type thing, okay. [UAM stands for uniform accelerated motion] Gravity is definitely in the mix, oh good, okay. So let's see first I'm going to draw a nice little picture cause it helps me think.

In order to segment this section of the protocol, I determined where the subject made transitions in thought. For example, after identifying the problem as a "UAM type thing" the subject moved to discussing gravity and then to delineating a start to the solution path. Therefore, I segmented this section as follows:

> **Segment one:** Okay so I'm thinking is uh… VF equals VI, one of those type of problems, um, UAM type thing, okay. [UAM stands for uniform accelerated motion]
>
> **Segment two:** Gravity is definitely in the mix, oh good, okay.
>
> **Segment three:** So let's see first I'm going to draw a nice little picture cause it helps me think.

The segmented audio was then coded using the verbal coding scheme that was developed based upon the premises of the data.

4.2.4.2 Coding Data

The coding scheme was a bit complex with up to four code levels. The main segments of the coding scheme were as follows:

- Metacognition
- Understanding or reading the problem
- Problem analysis
- Solution method

The entire coding scheme can be found in Appendix F.

The first level determinations were the easiest ones to complete and had high interrater agreement. An example of an understanding/reading comment is the following from M14:

> The rabbit gives chase the instant the turtle passes him as shown on the velocity time graph below. When do the two have the same velocity? [reading problem – stops at first question]. So if this is velocity time. [points to dotted line on graph]

The subject is holding the problem and reading it while interpreting the graph that came with the problem statement. Whereas, NM18 below is completing a different type of statement:

> Let's see here… Um… I realize I should put a table of contents in my notebook for days like this. Wait hold on what's this? That's not what I want either. Um… (*flipping pages*). Where's (*inaudible*) looking for... [looking through notebook but can't find what wants].

NM18 is completing a problem analysis task specifically in this case searching her notebook for an equation. Solving tasks were some of the easiest to determine as can be seen in the following excerpt from NM7:

> Equals negative big number 14 million. Okay and then we'll subtract by the quantity 50,000 times 3,000…29 million. Yeah. And then we divide by 10,000. Equals negative two hundred… no 29,000 meters per second. X-component of the asteroid. [finish inputting known values and solves for final velocity]

In this instance NM7 is obviously algebraically solving for final velocity.

The metacognitive statements involved comments where the participants were actively regulating their cognitive processes in order to solve the problem. Figure 4-1 shows the portion of the verbal coding scheme that would be used to code a statement after it had been determined that it was a metacognitive statement rather than an understanding the problem, problem analysis, or solving statement.

Metacognition: (actively monitoring and regulating cognitive processes used to solve the problem)

- Planning how to approach the problem
  - Setting goals
    - What type: check problem status, draw diagram, and search for equation….
  - Describe solution approach in words
  - "Remember" how to do this problem – (associated with problem type matching)
- Monitoring comprehension
  - General understanding of physics involved (self questioning, telling knowledge, Interpret graph (explain area, slope, make note))
  - General understanding of the problem
- Evaluating/checking (progress towards completing problem)
  - Approach/path/plan
    - Does it make sense?
    - Steps in path
    - Revise plan/model selection/problem type
  - Information used
  - Is equation appropriate?
  - Answer
    - Via alternate path
    - Recalculate
    - Does it make sense?
    - Units
    - Compare to problem - what was asked for (correct quantity?) or compare to given information
    - Compare to problem type example
    - Revise answer

Figure 4-1: Metacognitive segment of the verbal protocol coding scheme

After the determination that a statement was a metacognitive statement I would have to decide what type of metacognitive statement it contained. The following excerpt from NM7 is a classic metacognitive statement:

> Wait, no. I did something wrong here. Okay this is the force going down. Okay this is the force going down. The water stops the ball here… 5 meters per second. Okay… Um…

He realizes that he did something wrong and then starts interpreting his knowledge. This would be coded as a monitoring comprehension task. M30 made the following comment after she identified that one object in the problem was moving at a constant velocity:

> So they have the same velocity when the velocity of the rabbit is 5 so it's just a question of when these two lines intersect

In this instance M30 is planning how to solve the problem by actually describing the solution path in words. The following except is from M20 following her initial algebraic solution for problem 4 in the verbal protocol:

> Why didn't I… see that on the graph (*laugh*). It was on the graph; it was right there… It's easy, look there's a square there and then there's two half triangles there and then you know that there's two squares there…[showing area under curve by counting squares would have been easier].

This is an example of a metacognitive statement since she is monitoring the process of solving the problem. She is checking a previous response so it is also an evaluating/checking statement. Now I would need to decide whether the statement referred to checking the approach taken, information used, appropriateness of the equation, or answer. The subject in this case already has an answer and is now checking if that answer is correct; therefore, this statement would be coded as an evaluating answer statement. If the statement involved evaluating the answer the next step would be to decide if the subject was attempting to see if the answer made sense, check the answer via an alternate path, revise the answer, etc. Looking at the coding scheme it can be seen that M20 is checking the answer via an alternate path – in this case a graphical one. An example of a metacognitive statement where the subject, M14, is evaluating whether the answer makes sense as seen below:

> I'm not sure if that makes sense cause it's pretty high. It's really, really high. Hmm… So this one goes from positive… negative to positive since they hit each other. And it's really big; this one's small. And has negative velocity. And impacts with the smaller one, which has a high velocity coming this way. So this goes the opposite direction (*inaudible*) but this one goes this way and it would have to be… Okay, think that… Unless I had a calculator error I think that's right.

An entire segmented verbal protocol with codes, associates pictures and time stamps can be viewed in Appendix G.

In order to determine the number of physics errors and the number and type of metacognitive statements made by participants when on a productive path, I constructed physics protocol codes for each subject. This physics protocol code allowed me to delineate where and when the participants were on a productive physics path. I was able to follow the transitions in the physics path taken, discover physics errors made and the associated metacognitive statements produced for each subject. Figure 4-2 illustrates a section of one such physics protocol code. In this example a student is attempting to solve a problem contained in most physics textbooks that has one object accelerating past an object moving at a constant velocity and asks that the students determine where and at what position they will meet.

Rabbit:
$d = \frac{1}{2}(5/3)t^2$

valid equations; productive path

Turtle:
$= 5t$

$5t = \frac{1}{2}(5/3)t^2$ — **valid equations and productive path**

$t = 6$

And now they should have the same position so this is also a convenient way of checking it.
Metacognition > monitoring comprehension > general understanding of physics [15;10;14]

$D = (5m/s)(6s)$

$= 30$ m

And now for the rabbit D equals ½ 5/3 times 6 seconds squared. That's 5/3 meters per second squared and that should equal 7.5 times 5 divided by 3 times 6 squared. Mess that up… .5 times 5 divided by 3 times 6 squared. That also equals 30 meters. So they catch up right at that… So yes, so 6 seconds and 30 meters
Metacognition > evaluating/checking > answer>via alt path [15;50;25]

6 sec and 30 meters — CORRECT

Figure 4-2: Subject M-30 is solving a problem that requires her to determine when and where the turtle and rabbit meet. The highlighted statements are the metacognitive statements made by the subject while on the productive path.

Figure 4-3 illustrates a physics protocol code for a student who was initially on an unproductive path during a constant acceleration problem. The student initially invoked a valid equation but one not usable for solving this problem therefore it was determined that she was on an unproductive solution path. The student reassessed the problem, changed equations and then returned to the productive path.

$F = ma$ — **valid equation; non productive path**

I want to find the formula for (*inaudible*). [flipping through notebook again then finds something]
Metacognition > planning > setting goal > find formula [02;13;28]

$\Delta x = v_i\Delta t + \frac{1}{2} a\Delta t^2$ — **valid equation; productive path**

d= 442.2 m — CORRECT

Figure 4-3: Subject NM-19 is solving a problem that requires her to determine the height of a building from which a ball is dropped. The highlighted statements are the metacognitive statements made by the subject while on the productive path.

In the next problem NM-19 made an error during a momentum problem involving the collision of an asteroid and a comet. She incorrectly added the masses of the two objects together after the collisions. This was inappropriate and a statement indicating that a physics error was made and what the error was are listed on the code. However, she caught her error and then proceeded on her solution path. She then made a math error and did not successfully complete the problem and did not correct the final error. The physics protocol for this problem can be seen in Figure 4-4. An entire physics protocol code is contained in Appendix H.

$m_1v_1 + m_2v_2 = (m_1 + m_2)]$ **invalid equation; productive path**

[Erases the parentheses and the + m_2 on the after collision side]
Metacognition > evaluating/checking > is equation appropriate
Physics error > add masses together

↓

$m_1v_1 + m_2v_2 = m_1v_1' + m_2v_2'$ **valid equation; productive path**

That seems really… large
Metacognition > evaluating/checking > answer > does that make sense

↓

$v_a' = 254{,}000$ m/s INCORRECT - MATH ERROR
Does not catch math error

Figure 4-4: Subject NM-19 is solving a problem that requires her to determine the final x-component of an asteroid's velocity after it has collided with a comet. The highlighted statements are the metacognitive statements made by the subject while on the productive path.

4.2.4.3 Correlation between Study 1 and Study 2

All of the student volunteers in this study also participated in Study 1; therefore, it awarded the opportunity in this thesis to determine if their problem-solving and metacognitive behaviors could be predicted by the task scores taken in the previous study. In order to determine the predictive ability of Study 1's paper and pencil tasks the proportion of problem-solving behaviors observed in Study 2 were correlated with the scores from Study 1.

4.3 Hypotheses and Predictions:

Study 1 showed that modeling participants scored higher than non-modeling participants on both the FCI and the PS Task. In addition, Study 1 demonstrated that the modeling participants constructed more "expert-like" knowledge structures during the course of a single school year. However, Study 1 was unable to test for differences in problem-solving strategies utilized by the modeling and non-modeling participants. Study 2 was designed such that it would allow for any differences in problem-solving strategies between the two groups to come too light. Based upon on my understanding of the literature and the results from Study 1, I predict the following:

- It is expected that the modeling participants' performance on the problems designed for Study 2 would demonstrate a greater percentage of correct answers with fewer physics errors committed.
-

This prediction seems reasonable given that in Study 1 the modelers scored significantly higher on both the Force Concept Inventory and the Problem-Solving Task. Modeling students outperformed the non-modelers on the problem-solving task designed for Study 1; thereby, demonstrating enhanced problem-solving skills.

- I would expect the non-modeling participants to display more “novice-like” problem-solving strategies while the modeling participants should demonstrate more “expert-like” problem-solving strategies.

The above prediction seems reasonable since Study 1 demonstrated that the non-modeling participants have a significantly lower “expert-like” score than the modeling participants. This finding suggests that the non-modeling participants have not constructed an “expert-like” knowledge structure. Therefore, given the findings in past literature described in Chapter 2 one would expect the non-modeling participants to display more “novice-like” problem-solving strategies. The non-modeling participants would be expected to take longer to solve a problem, spend a greater amount of time referring to the problem statements, have more reliance on equation lists, classify problems by problem type more often due to a reliance on surface features, demonstrate less flexibility in solution methods (i.e., choose mostly algebraic solutions rather than graphical solutions), work backwards towards the solution and initially identify an equation needed to a solve each problem rather than a model or concept. A comprehensive list of novice-expert differences can be found in Table 2-1. All of these behaviors can be tested for via the verbal protocol study and the coded statements in the area of problem analysis, understand/read the problem, and solution approaches as described in Section 4.2.4.2. By analyzing the proportion of statements made we should be able to determine if the students are behaving more expert-like or novice-like.

- The modeling participants should spend a greater time considering the problems on a metacognitive level because of the modeling pedagogy.

Metacognitive abilities should be enhanced in the modeling participants as it is implicit in the modeling method that one clarifies choices made during problem-solving steps. This ability is constantly practiced by the modeling participants as they are routinely asked to defend what they know and how they know it. In addition, the research described in Chapter 2 demonstrated that the expert-like subjects performed more metacognitive tasks such as evaluating the solution path to a greater extent. Therefore, this instructional strategy should train the modeling participants to consider their problem-solving solutions metacognitively to a greater extent than non-modeling participants. However, if my prediction that the non-modelers would have less success obtaining the correct answer is accurate one might suspect that the non-modeling group could demonstrate more metacognitive activity as they struggle with their solution approach. However, if this is the case then the metacognitive activity across the entire solution path of the modelers and non-modelers would at the minimum demonstrate equality in metacognitive activity. This equality in the amount of metacognitive activity would occur because the modelers will be trained to think metacognitively about the problem-solving strategy both when they are on a productive solution path and when they are not. Therefore, over the entire solution path the proportion of metacognitive activity between the two groups should at a minimum be equal or greater for the modeling participants.

- If one looks at the sections in which the modelers and non-modelers are both on a productive solution path I would predict that the modeling participants would not only demonstrate greater metacognitive activity in general but would also produce greater amounts of evaluation/checking statements. On the other hand, while the non-modelers might be producing fewer metacognitive statements while on a productive path they will be producing a greater proportion of monitoring comprehension statements rather than planning or evaluation statements. Therefore, the two groups' metacognitive statements will serve a different purpose.

A productive path is considered to be a physics path that could lead to a correct solution approach. When on a productive path, after selecting a model to guide their solution the modelers should be deploying their model without need to reflect upon the correctness of the model chosen. However, the modeling pedagogy is always asking the modelers to make sense of and defend their solution paths and their final answer therefore it makes sense that the modelers would evaluate their solution path to a greater extent than the non-modelers. In the case of the non-modeling participants the prediction that they would be making a greater proportion of monitoring comprehension statements than the modelers might seem counter-intuitive. However, after the model selection the modelers would be deploying their model without need to reflect on if the model is correct therefore they should not feel a need to monitor their comprehension once they have deployed a model. This is in line with the research in Chapter 2 that had found that expert-like subjects consistently evaluate their solution strategy and once deciding on a physical principle they very quickly have solution procedures at their disposal so they simply run the model to determine the answer.

- When looking at the specific types of evaluation statements made by the two groups I would predict that as the modelers evaluate the solution path they will evaluate the approach or the answer to a greater extent than the information used or the equation chosen. On the other hand, I would expect that when the non-modelers do evaluate their solution path they would do so via the information used and the equation chosen to a greater extent. For example, the non-modelers might evaluate the information used with the problem statement or check to make sure that they selected the correct equation from their equation list. However, the modelers would be evaluating their approach and answer to a greater extent since they should be asking themselves why they are using the approach they are and if the answer/approach make sense.

This prediction would be in line with the "novice-like" problem-solving strategies that the non-modeling participants would be expected to demonstrate. The non-modelers will rely more upon the equation and how the information contained in the problem statement plugs into the equation selected as described in Table 2-1. Then since the modelers are constantly reminded to defend the choice of approach or model and look to see if the answer is in line with what they would expect from the model they should not lean towards being overly dependent upon the equations or information contained in the problem.

- When reviewing the findings at the next level of metacognition, I would predict there are differences between how the two groups evaluate the answer and the approach. I predict that the modelers would be attempting to make sense of their answer and approach to a greater extent than the non-modelers as well as evaluating the answer more often via alternate approaches.

This prediction makes sense since it is an implicit part of the modeling pedagogy that the students are asked and must demonstrate that their problem-solving steps as well as their answers make sense in light of the model chosen. In addition, the modelers should be evaluating their answer via alternate approaches more often since the representations (i.e., verbal, diagrammatic, algebraic and graphical representations) contained in each model should allow the modelers to be more flexible with a variety of strategies with which to solve problems once they deploy a model.

- It is expected that while the modeling participants will commit fewer physics errors that they will also be able to catch and correct a greater percentage of physics errors than their non-modeling counterparts.

This prediction should occur if the previous prediction concerning enhanced metacognitive activity on the part of the modeling participants is realized. If the modeling participants are evaluating their approaches and answers to a greater extent than the non-modelers this should allow the modeling participants to more easily recognize any physics errors that are produced.

- It is expected that the participants' expert-likeness as demonstrated by their FCI, PS Task, and card sort scores from study 1 will correlate highly with the problem-solving behavior observed via the verbal protocol study.

If a subject in study 2 had a high expert, low surface feature or high FCI score then I would suspect that the subject would demonstrate behaviors that are more on par with the expert-like problem-solving behaviors listed in Table 2-1. However, subjects that had low expert, high surface feature or low FCI scores should demonstrate the novice-like behaviors listed in Table 2-1. If this reasoning is correct than correlations between these scores and problem-solving behaviors observed in the verbal protocols should support this prediction.

Overall, the predictions as related to the research questions are that the modeling participants will demonstrate enhanced problem-solving strategies. These enhanced problem-solving strategies should allow for more "expert-like" performance allowing for greater error recovery.

4.4 Results

All of the quantitative results noted in this section include only the talk aloud protocol data and none of the data from the retrospective interview. When the retrospective data was included in the analysis the proportions discussed below increased in value.

4.4.1 Coding and Reliability

All twenty-eight transcriptions were coded using the verbal coding scheme in Appendix F. A second rater coded a random 30% of the participants after being trained to use the verbal coding scheme. These codes were checked for interrater reliability on all four coding levels using Cohen's Kappa (a non-parametric correlation). Table 4-1 summarizes the correlation results not including the retrospective protocol data. When the retrospective data was included in this analysis the correlation values decreased slightly.

TABLE 4-1: Cohen's Kappa Correlation Findings for Interrater Reliability Check

Code Level	K-value
First	0.95
Second	0.90
Third	0.85
Fourth	0.87

These results were quite satisfactory and demonstrated excellent agreement between the two raters, especially at the first two levels of the code. Any coding differences were resolved via discussion and then any changes were applied to all the coded transcripts.

4.4.2 Solvers vs. Non-Solvers and Time Taken on Solution Paths

The data was analyzed for correctness and the time taken for each attempted solution. The data showed that sixty-eight percent of the verbal protocol problems solved by modelers were correct whereas only forty-eight percent of the solutions produced by non-modelers were correct. The Chi-square 2 X 2 table of results (Table 4-2) demonstrates that this is significant difference.

TABLE 4-2 Solvers vs. Non-Solvers

	Modelers	Non-Modelers	Total
Correct Solution	91	37	128
Incorrect Solution	42	40	82
Total	133	77	210

(Chi-square test = 8.50, $p < 0.01$, df = 1)

Table 4-3 shows the time devoted by each subject to each problem (correct answers are highlighted). Note that all of the participants solved problem 4a correctly and all of the non-modelers solved problem 1 correctly.

The times in Table 4-3 were tested for significance utilizing a one-tailed t-test. The t-test is typically robust against violations in variance when the sample sizes are similar. However, the sample sizes in this study are different therefore I completed a Levene's test for equality of variances in order to determine the degrees of freedom and the p-value for all the results that follow in this study. Table 4-4 summarizes the statistics for solution times.

The differences in the time taken to solve the problems were significant for only problems 4b, 4c, and 5 (p-values ranging from 0.02 to 0.0005). Problem 2 is marginally significant with a $p < 0.06$ in favor of a quicker solution time for the modeling students. It is important to note that all of the participants solved Problem 4a correctly demonstrating that all of the participants could utilize and solve problems using graphical representations and the difference in time is not significant between the two groups. In the case of Problem 3 the solution times were very similar. The non-modelers demonstrated mastery of Problem 1 since they all solved it correctly while the modelers did not and therefore the solution time is shorter for them although not statistically significant.

TABLE 4-3: Summary of Time Taken (in seconds) and Correctness of Verbal Protocol Problems

Modeling participants are designated by an M whereas the non-modeling participants are designated by an NM (the bottom shaded portion of the table). If the problem is correct the time is highlighted in blue.

Subject	Problem 1	Problem 2	Problem 3	Problem 4a	Problem 4b	Problem 4c	Problem 5
M1	334.2	241.4	394.5	50.5	91.6	84.1	400.2
M3	235.8	125.9	326.4	52.6	145.5	56	132.2
M4	218.2	121.9	358.4	48.2	49.7	201.9	237.9
M5	141.3	150.9	339.1	67.2	60.7	122.1	277.5
M6	187.3	132.7	359.3	108.1	38.1	32.9	306.8
M13	102.8	72.5	314.7	48.1	84.4	83.3	128.8
M14	239.7	76.3	313.3	58.7	33.4	196.5	684.1
M16	105.1	408.1	456	56.2	34.3	63	568.9
M20	180.6	129	345.5	94.3	37.8	407.5	84.8
M23	126	186	250.7	45.7	23.9	86.6	253.2
M24	118.9	71.6	214.8	36.4	44.8	106.3	170.2
M26	86.8	189.4	569.6	36.7	34.6	372.7	91.3
M29	77	91.6	289.9	52.7	24.8	123.7	137
M30	138.2	151.9	313	39.8	89.5	119.3	173.1
M32	84.6	118.7	184.2	29.6	30.2	152.3	97.6
M41	307.5	155.8	756	120.9	268.8	38	320.2
M45	701.1	435.1	451.9	53.1	51.5	271.4	271.3
M48	235.5	299.9	409.9	74.2	66.5	84.1	237.7
M61	530.8	216.7	638.1	147.7	297.1	255.5	775.1
AVG M	218.5	177.2	383.4	64.2	79.3	150.4	281.5
NM2	128.1	113.2	297.9	37.5	38.3	265.1	821.9
NM4	258	127.1	741.7	646.5	250.2	436.7	616.4
NM7	91.5	407.5	358.3	48.8	67.5	423.8	253.2
NM11	131.4	488.6	634.3	38.3	77.1	137.8	684.5
NM16	258.6	131.8	321.3	53.3	84.3	266.5	250
NM18	118.7	167.3	293.4	49	154.7	491.3	404.6
NM19	173.5	711.4	228.3	67.2	81.1	419.3	382.1
NM20	295.1	368.7	272.2	44.2	298	457.3	344.3
NM21	253.9	242.2	424.4	55.3	198	499.3	513.8
NM22	296.8	236.8	158.2	95.7	123.3	101.6	648.4
NM23	148.4	121.1	520.8	70.9	232.2	240.6	368.7
AVG NM	195.8	283.2	386.4	109.7	145.9	340.0	480.7

TABLE 4-4 Summary of Statistics for solution times
Problems that are statistically significant are highlighted in pink

Problem Number	M Avg. Time	NM Avg. Time	Levene's test (sig.)	One-Tail Prediction	p-value	df
1	218.5	195.8	0.248	M>NM	0.33	28
2	177.5	283.2	0.021	M<NM	0.057	13.5
3	383.4	386.4	0.324	M<NM	0.48	28
4a	64.2	109.7	0.037	M<NM	0.21	10.4
4b	79.3	145.9	0.293	M<NM	0.0195	28
4c	150.4	340.0	0.093	M<NM	0.0005	28
5	281.5	480.7	0.716	M<NM	0.006	28

4.4.3 Verbal Protocol Coding Scheme Analysis

4.4.3.1 Comparison of the Proportion of Understanding, Problem Analysis, Solving, and Metacognitive Statements

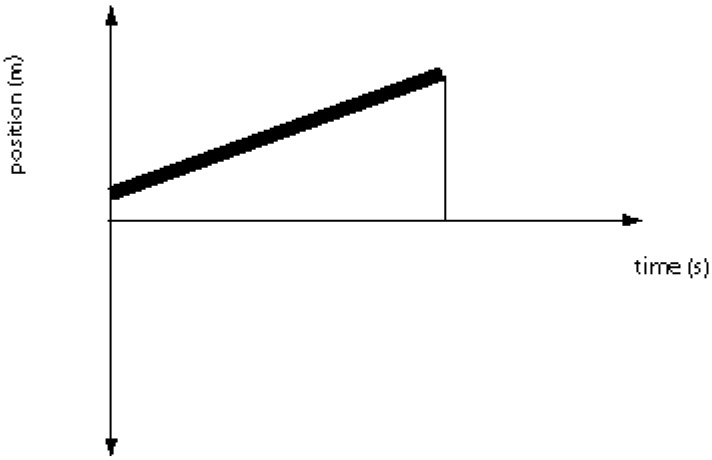
The first step in the verbal protocol analysis was to see if the groups had differing proportions in any of the main levels of the code. An analysis of the proportion of the first-level statements vs. the total number of statements illustrates that both groups of participants had a similar proportion of problem analysis, solving, and metacognitive statements which were not significantly different. However, the two groups of participants differed in the proportion of understand/read statements to total statements as predicted. The non-modelers demonstrated a proportion of 25% whereas the modelers had a proportion of 21%. Levene's test for equality of variances showed that the variance in this case was not significantly different. A t-test showed that the difference in the proportion of understand/read statements was significant between the two groups (NM>M, $t_{28} = -2.399$, $p<0.012$).

4.4.3.2 Comparison of the Proportion of Graphical and Diagramming statements

The models developed by the modeling participants contain specific graphical and diagrammatic representations. For example, when constructing a model of constant velocity the modeling subjects distinguish between four different representations that all describe constant velocity. Upon first developing the model the modelers would describe the model verbally by stating how the position changes with time. For example, they might say that the change in position is the same for each change in time. The modelers would specify the following algebraic representation (which the non-modelers would simply know of as an equation) for this model:

$$x_f = v\Delta t + x_i$$

The modelers would also specify that the above algebraic representation was derived from and described in the following position vs. time graph:



Finally, the modelers would specify a diagrammatic representation of constant velocity such as a motion map. An example of this type of diagram would be as follows:

This diagram illustrates that the object is covering the same position each time segment as the length between each arrow is constant. The length of the arrows demonstrates the constant velocity since they are all the same length. Therefore, since the modelers work consistently with these different graphical and diagrammatic representations it was predicted that the modelers would produce a greater proportion of statements referring to the use of graphs and diagrams as they would deploy these representations when using a model to guide their solution path. In order to determine the proportion of these statements, I collapsed all of the graphical and diagrammatical statements made within the understand/read, problem analysis and solving categories. The differences in these proportions were tested for statistical significance between the two groups. The modelers were found to have a graphical proportion of 9% vs. the non-modelers proportion of 5%. Levene's test for equality of variance demonstrated that the variance could not be assumed to be equal for the graphical proportion ($p = 0.012$). A t-test, assuming unequal variances, illustrated that the difference in the proportion of graphical problem statements was significant between the two groups (M>NM, $t_{23.} = 6.29$, $p<0.0005$). The diagrammatical proportion of 3% for modelers and 4.6% for non-modelers was not statistically significant and was not in the direction predicted. It is possible that in order to observe a difference one would need to code for the quality of diagram used since in this study any diagram be it a sketch of the situation or a motion map was coded as a diagram.

4.4.3.3 Comparison of the Proportion of Textbook or Notebook Searches vs. other Problem Analysis Tasks

The non-modelers searched their textbooks, notebooks and equation lists to a greater extent that the modelers. The proportion of problem analysis tasks either searching their textbooks or notebooks was 11% for the non-modelers and 2.2% for the modelers. A Levene's test of equality of variance showed that the variance could not be assumed to be equal ($p = 0.001$). A t-test, assuming unequal variances, demonstrated that the difference in the proportion of searching was significant between the two groups (NM>M, $t_{13.43} = -2.920$, $p<0.006$) as predicted.

4.4.3.4 Comparison of Problem Classification Codes

The coding scheme allowed for a distinction between problem classifications based upon identifying models vs. problem types. When participants were classified as identifying a model they did not have to use the name of the model since the non-modelers would not have been trained in the specific model names. Rather, identifying a model was interpreted as identifying the concept that was associated with the problem. For example, in the following excerpt M5 has finished reading a problem and says:

> All right so we need to find the distance. So we need to figure out the model that we can use for mass and time to figure out the distance and it's going to be accelerating 9.8 meters per second so… I always forget the constant acceleration model…

M5 is identifying that he wants to deploy the constant acceleration model. This segment would be coded as identifying a model. In the following excerpt NM2 has finished reading problem 5 dealing with types of slides and he says:

> Um… Okay… So let's see… I think its energy.

While NM2 does not use the term energy model he is identifying the concept that must be deployed in order to solve this particular problem. This section of NM2's verbal protocol would be coded as identifying a model.

The following excerpt from NM18's verbal protocol shows a distinct difference from the two above. NM18 had been reading problem 2 and immediately after reading the word collision in the problem he said:

> Okay great. Um… I know I have a problem like this in here somewhere… Now… Um… well, let's see… [starts to look through notebook then gives up on it]……

NM18 then went in search of the similar problem type in his notebook. Therefore, he would not be coded as identifying a model but as identifying a problem type. As illustrated, identifying a problem type is recognizing that this particular problem was similar to some past problem solved by the subject via a surface feature within the problem.

The coding scheme allows one to determine if the participants were identifying equations or algebraic representations linked to a model. Participants were coded as identifying an equation if the equation selected came after the subject had previously mentioned looking for an equation, searched an equation list or identified a problem type. For example, NM18 above indicated that he was looking for a problem type at which point he started to look through his notebook and then he said:

> Okay um… (*whispering*) F Net equals… I'm thinking maybe this would be it. I'm thinking maybe I want F Net equals MA [writes F_{net}=ma]

In this instance this segment would be coded as identifying an equation. A student was coded as identifying of an algebraic representation only if they had initially identified the model that pertained to that problem within the last two statements. After reading the problem that was designed to use the constant acceleration model M5 said:

So we need to figure out the model that we can use….. and it's going to be accelerating 9.8 meters per second so…the constant acceleration model

Immediately after M5 identified that he was going to deploy the constant acceleration model within one statement he said:

I know it's like… Wait we don't even know…[writes $x_f= 1/2at^2$]
Well… $x_f = ½$… .5… Um… plus zero so it's… That should be it right there.
[fills in the equation with known values].

This segment would have been coded as identifying an algebraic representation as this is one of two algebraic representations contained in the constant acceleration model. In addition, this segment demonstrates that the algebraic representations seems to be chunked with the model since immediately after identifying a model the student within one statement stated the algebraic representation attached to that model. This is very similar to that of the expert/novice studies described in Chapter 2. There were a number of these instances observed in the modeling students' verbal protocols. In a number of cases I could not distinguish between the two possibilities and the statement was coded as could not distinguish. Due to the strictness of the coding definition the proportion of could not distinguish was quite high and similar in number for both groups.

Any equation or algebraic representation was then coded as either forward chaining or working backwards. A strict definition was adhered to in this case. A subject was considered to be working backwards if an equation or algebraic representation selected contained the variable that was asked for in the problem statement or followed from the selection of a problem type. The only exception to this was when the algebraic representation directly followed the declaration of the model to be used in which case it could contain the unknown variable. For example, the excerpt above from NM18 would have been coded as working backwards since the equation mentioned contains the unknown variable and it was preceded with the declaration of identifying a problem type. The excerpt from M5 would have been coded as forward chaining since it directly followed the declaration of the model to be used. However, directly following the previous excerpt by NM2 he says:

…let's see… Um… mgy equals ½ mv squared. [writes $mgy = ½ mv^2$].

This segment would have been coded as not only identifying an algebraic representation but also forward chaining.

The differences in the proportion of problem classifications were statistically significant in all areas except the "cannot tell" category. Table 4-5 summarizes these findings.

Table 4-5: Summary of Statistics for Problem Classification

Categories that are statistically significant are highlighted.

Problem Classification Category	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Id Model	33	16	0.271	M>NM	2.944	0.003	28
Id Problem Type	0.7	8	0	NM > M	-3.079	0.005	12
Id equation	18	45	0.701	NM > M	-3.409	0.001	28
Id Algebraic representation	28	9	0.205	M > NM	3.881	0.0005	28
Can't tell	21	17	0.568	None	0.548	0.59	28

As can be seen in Table 4-5 the statistics are in line with the initial predictions. As predicted the modelers identify the model associated with the problem significantly more that the non-modelers (33% to 16% ($p<0.003$)) demonstrating a reliance on deep structure (i.e., models). In addition, the non-modelers are more likely to identify the problem type associated with the problems showing a reliance on surface features. Equal variances could not be assumed due to the significance of Levene's test but the differences were still highly significant ($p < 0.005$) for the proportion of identifying problem type statements. The non-modelers identify an equation to solve the problem 45% of the time whereas the modelers demonstrate this identification only 18% of the time ($p < 0.001$). On the other hand, the modelers identify an algebraic relationship 28% of the time vs. the 9% demonstrated by non-modelers ($p< 0.0005$). The implications for these approaches are tremendous. By identifying problems based upon the models used to solve them the modelers are demonstrating the use of deep structure and a more coherent knowledge structure. The modeling students should have an easy search space since they only have six models to distinguish between. After they select the model via a breadth search across all of the models all of the representations (algebraic, graphical, diagrammatic, and verbal) are chunked with the model and immediately become available for the students allowing them to be more efficient (i.e., should take less time to reach a solution) and demonstrate more flexibility. The non-modelers identification of problem types and equation lists is not an efficient strategy and may have caused the increase in solution times observed in Section 4.4.2. The non-modelers must do a depth search though a large list of equations to find the appropriate one or they can look through a number of problem type examples attempting to uncover one that has the same surface feature and can be used as a guide during the solution.

The last level of the problem classification distinguishes between whether the participants solve the problems using forward chaining or working backwards methods. I initially predicted that the non-modelers would work backwards whereas the modelers would use forward chaining to a greater extent. Table 4-6 summarizes the findings.

TABLE 4-6: Summary of Working Backwards vs. Forward Chaining

Categories that are statistically significant are highlighted.

Forward chaining vs. working backwards	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Working backwards selections	26	52	0.214	NM > M	-3.572	0.0005	28
Forward changing selections	70	47	0.325	M > NM	2.901	0.0035	28

As per Levene's test, as shown in the table, equal variances can be assumed in both cases. The table shows that the modelers worked in a forward chaining manner to a proportionately greater extent with 70% of the proportion of the statements coded as forward chaining vs. 26% coded as working backwards. On the other hand, the non-modelers worked backwards to a greater proportion (52%). These findings were in line with the initial predictions and highly significant with p-values of 0.0035 and 0.0005.

4.4.3.5 Solving Code Analysis

It was initially predicted that the modelers would choose to complete more graphical problem solutions while the non-modelers would be more algebraic in their solution path. The four solving codes were: algebraic solution, graphical solution, guess and check and cannot distinguish. Table 4-7 summarizes the findings for this section of the protocol analysis.

TABLE 4-7: Summary Solving Protocol

Categories that are statistically significant are highlighted.

Solving Code Proportion	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Algebraic Solution	62	83	0.158	NM > M	-4.351	0.0005	28
Graphical Solution	28	7	0.006	M > NM	7.031	0.0005	21.4
Guess and Check	7	4	0.47	None	0.842	0.41	28
Cannot Distinguish	2	2	0.158	none	0.230	0.82	28

The Levene's test revealed that only the graphical solution method had unequal variances. The t-test, assuming equal variances, shows that non-modelers produce algebraic solutions to a significantly greater extent than the modelers (83 % vs. 62%, $p < 0.0005$). The graphical solutions are even more telling with the modelers producing a significantly greater number of these solutions (28% vs. 7%, $p < 0.0005$). The proportion of guess and check solution methods were not significant between the two groups.

The type of solution method chosen by the participants utilizing a guess and check solution strategy was analyzed to determine if there were a difference in the type of strategy chosen (i.e., graphical vs. algebraic). Of the participants who chose to complete guess and check solutions 89% of the modelers utilized a graphical strategy vs. no graphical strategy usage by the non-modelers. This finding was statistically significant, assuming equal variances (M>NM, $t_{10} = 4.472$, $p<0.0005$). The non-modelers produced 100% algebraic guess and check solutions while the modelers only chose this path 11% of the time. This finding was also statistically significant, assuming equal variances (NM>M, $t_{10} = -4.472$, $p<0.0005$). These results highlight the greater flexibility of the modeling students in terms of solution paths open to them.

4.4.4 Analysis of Metacognitive Statements on a Productive Path

It has been observed that the two groups do not show any difference in the proportion of metacognitive statements over the course of the entire protocol. However, it was predicted that the proportion of metacognitive statements would be significantly higher for the modelers when comparing only those statements made when the participants are on productive paths. The physics codes were used to determine when the participants were on a productive physics solution path. These areas were then correlated with the verbal code in order to obtain data to determine the proportion of metacognitive statements and the types of metacognitive statements produced. The modelers metacognitive statement proportion of 27% when on a productive path vs. 21% for non-modelers proved to be statistically significant (M>NM, $t_{28} = 2.009$, $p<0.03$).

The next step is to look at the proportion of types of metacognitive statements made while on the productive path. The findings below are in line with my predictions. Thirty-nine percent of all metacognitive statements made by non-modelers were in the form of monitoring comprehension; whereas, only twenty-four percent of the statements produced by modelers were in this area. The difference in the proportion of monitoring comprehension statements was found to be significant between the two groups (NM>M, $t_{28} = -2.789$, $p<0.005$). On the other hand forty-three percent of all of the metacognitive statements made by the modelers while on a productive path were in the area of evaluating/checking. Only twenty-nine percent of the non-modelers metacognitive statements were evaluative in nature. The difference in the proportion of evaluating/checking statements between the two groups was found to be significant (M>NM, $t_{28} = 2.131$, $p<0.02$).

The next level of analysis is to determine the differences in proportion for the four different types of evaluating/checking statements. The participants' evaluating/checking statements were coded into four categories. The participants evaluated the appropriateness of the equation, the information used, the approach, or the answer. Table 4-8 summarizes the findings in this area.

TABLE 4-8 Summary of the Type of Evaluating/Checking Statements.

Categories that are statistically significant are highlighted.

Evaluating/checking Statement Proportion	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Is equation appropriate?	7	18	0.074	NM > M	-1.501	0.07	28
Information Used	6	17	0.048	NM > M	-1..216	0.12	11.2
Approach	33	31	0.247	M > NM	0.188	0.43	28
Answer	54	33	0.028	M > NM	1.830	0.0425	16.6

As predicted the majority of the modelers evaluating statements are dealing with the answer and the approach to the problem. The modelers seemed to check the answer 54% of the time whereas they checked the approach only 33% of the time. However, the non-modelers also evaluated these areas more often than the other two areas which was not predicted. The difference in the proportion of checking the answer between the two groups was statistically significant whereas both the proportion of checking the approach was not significant between the two groups. The non-modelers checked the equation to a greater extent at a marginally significant level which was as predicted. However, the proportion of evaluations concerning the information used was not significantly different between the two groups.

If one collapses the two top evaluative choices, answer and approach, it is found that the modelers utilize these evaluative tools 87% of the time vs. 69% for the non-modelers. Assuming unequal variances these differences are statistically significant (M>NM, $t_{28} = 2.14$, $p<0.047$). In a similar vein if one looks at the proportion of statements non-modelers made that evaluated either the information used or the equation one finds that they did this 31% of the time whereas, the modelers did this only 13% of the time. This difference is statistically significant assuming unequal variances (M>NM, $t_{28} = -2.14$, $p<0.047$).

When evaluating the approach taken the participants seem to do one of three tasks: they evaluate the steps in the path, they revise the approach, or they see if the approach makes sense. When the modelers are evaluating the approach taken, the majority of the time they are revising their approach (54%). The proportion of statements in which modelers seem to question if the approach makes sense and if the steps in the approach are correct are equivalent (19%). The non-modelers seem to revise the approach by a large amount (73%) maybe hinting at the lack of confidence that they have on a productive path. The proportion of statements in which the non-modelers evaluated the steps in the approach was 27% while they never seemed to check if the approach made sense to them. The only type of evaluating approach statement that proved to be significantly different between the two groups was the proportion of checking completed to determine if the approach made sense. Table 4-9 summaries the findings in this area.

TABLE 4-9 Summary of the Type of Evaluating the Approach Tasks completed by Modelers and Non-Modelers

Categories that are statistically significant are highlighted.

Evaluating Approach Task Proportion	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Does approach make sense	19	0	0.002	M > NM	3.004	0.004	17
Revise approach	54	73	0.831		0.831	0.296	23
Evaluate steps in approach	19	27	0.636		-0.443	0.662	23
Cannot distinguish	7	0	0.017		1.761	0.096	17

One can now ask: what in particular about the answer are the groups evaluating? The protocol delineated eight different answer evaluations that seem to take place. Table 4-10 summarizes the findings for these eight types of answer evaluations.

TABLE 4-10 Summary of the Type of Evaluating the Answer Statements made by Modelers and Non-Modelers

Categories that are statistically significant are highlighted.

Evaluating Answer Statement Proportion	M Avg. (%)	NM Avg. (%)	Levene's Test (sig.)	One-Tail Prediction	T Statistic	p-value	df
Via alternate path	23	0	0.008	M > NM	3.692	0.001	17
Recalculate answer	22	26	0.815		-0.237	0.815	22
Does the answer make sense	34	26	0.381	M > NM	0.562	0.29	22
Are the units correct	2	6	0.09		-1.027	0.316	22
Compare answer to problem givens	5	11	0.013	NM .> M	-0.806	0.22	6.1
Compare answer to problem type	0	12	0	NM > M	-1.853	0.06	5
Revise answer	3	4	0.774		-0.185	0.855	22
Cannot distinguish	11	11	0.996		-0.031	0.976	22

The only proportion which is significant is that the modelers evaluate the answer via an alternate path more significantly than non-modelers (23% vs. 0%, respectively). This result really demonstrates that the modelers are much more flexible in their solution approach and the chunking of the representations with the model probably allows this to occur. Table 4-10 shows that the modelers do attempt to make sense of the answer to a greater extent than the non-modelers (34% vs. 26%) although not to a significantly different amount. An interesting aside is that when one looks closely at the sense making one finds that 40% of the sense making done by the modelers occurs when they are checking if the sign of the answer makes sense; whereas, the non-modelers spend only 20% of their time on this subset of the task. These findings are in the same direction as the initial predictions. Both groups have the same proportion for recalculating their answers. The findings show that the non-modelers compare the answer to the problem givens and the problem type to a greater degree than the modelers; however, not significantly. When the non-modelers would compare an answer to a problem type they would actually do something similar to when they identified a particular problem type (i.e., keying into surface features). The non-modelers would go back to the text or notebook and review the answer by comparing it to the answer found in their problem type sample.

The above analysis deals with the proportion of the responses and does not look at how many and what percent of each group actually undertake to evaluate their answer along these dimensions (at some time during the problem-solving session). In the case of comparing the answer to a problem type, three out of eleven of the non-modelers did this while none of the modelers checked their answer via this technique (30% vs. 0%, respectively). This difference in percentage seems to be highly unlikely to not be a significant difference between the two populations. The following 2 X 2 table (Table 4-11) summarizes the number of participants who did and did not evaluate the answer by comparing it to a problem type. The chi-square test is significant.

TABLE 4-11 Evaluating Answers by Comparing to Problem Type

	Modelers	Non-Modelers	Total
Compare to problem type	0	3	3
Do not compare to problem type	19	8	25
Total	19	11	28

(Chi-square test = 5.8, $p < 0.025$, df = 1)

If one undertakes the same analysis for sense-making one discovers that thirteen out of nineteen modelers (68%) check to see if the answer made sense when only three out of eleven non-modelers (23%) did this. The following 2 X 2 table (table 4-12) summarizes these significant findings.

TABLE 4-12 Evaluating answers by Sense-Making

	Modelers	Non-Modelers	Total
Does answer make sense	13	3	16
Does not do any sense-making	6	8	14
Total	19	11	30

(Chi-square test = 4.7, $p < 0.05$, df = 1)

This shows that a much larger proportion of modelers (68%) evaluate their answers while on the productive path by considering if the answer makes sense, compared to 27 percent for non-modelers.

4.4.5 Analysis of Physics Errors

Utilizing the physics codes, the number of physics errors committed, which of these were discovered and which were corrected were determined for each subject. The number of physics errors committed by each group differed significantly. The modeling group on average committed 1.95 errors while the non-modeling group on average committed 2.91 errors (NM > M, $t_{28} = -2.24$, $p<0.017$). Not only did the modelers make fewer errors, they discovered and corrected the errors to a greater percentage. Levene's test for equality of variance determined that equal variances could not be assumed in the case of the correction rates. However, the modelers had a significantly greater correction rate of 23.1%; while, the non-modelers correction rate was 5.3% (M > NM, $t_{22} = 2.09$, $p<0.024$). An exploration of the physics codes was conducted to determine the differences and similarities in physics errors produced and caught between the modeling and non-modeling groups.

4.4.5.1 Types of Physics Errors Produced

There were a total of eleven types of errors produced by the participants. There are several commonalties in the types of errors. The same percentage of participants in the non-modeling and modeling groups ignored forces on objects ($\chi^2 = 0.072$, df=1, $p<0.789$) and confused the final velocity with average velocity during solutions ($\chi^2 = 0.639$, df=1, $p<0.424$). For example, in one problem the majority of the students forgot to include the force of gravity in the calculation of the net force. This is a common error among students. In several problems, the participants would calculate the final velocity and then substitute this value in for the average velocity when determining the position of an object.

Two types of errors were committed to a significantly different amount by each group. These errors included: confusing a velocity time graph for a position time graph and using variables from different dimensions in the same equation. For example, in one problem students would routinely utilize the acceleration due to gravity which is in the y-dimension in the same equation with a horizontal change in position thus committing the error of utilizing variables from different dimensions in the same equation. Ninety percent of all non-modelers committed this error while only seventy percent of modelers committed it ($\chi^2 = 4.59$, $p<0.032$, df=1). In one particular problem participants routinely confused a velocity time graph for a position time graph thus reading off position data when it was actually velocity data. This confusion between velocity vs. time graphs and position vs. time graphs was quite notable between the two groups and statistically significant. While only 20% of all modelers committed this error, 50% of the non-modelers committed the same error ($\chi^2 = 4.98$, $p<0.026$, df=1).

4.4.5.2 Types of Physics Errors Caught

The types of physics errors that are caught are very telling. It seems that once a subject committed the error of using variables from different dimensions in the same equation there was no turning back since not a single subject corrected this error in either group. In addition, ignoring forces seems to have been an error that is difficult to correct since none of the participants did so. The subjects as a group, both non-modeling and modeling, seemed to believe that when finding a force all they needed to do was to multiple the mass times the acceleration to determine the amount of the force in question without taking into account other forces such as a normal force or the force of gravity on the object. It seems that the students would greatly improve their performance if they simply drew force diagrams as retrospectively many of the students, especially the modelers, realized their error as soon as a detailed question was asked. One student even said that he felt he did not need to draw a force diagram on paper since he could do one in his head.

Both groups had equal success in catching minor errors such as adding masses together when they should be considered separately, using the wrong acceleration rate, and ignoring signs. The modeling group enjoyed great success catching several types of errors although not to a statistically significant amount. Seventy percent of the errors confusing final velocity with the average velocity and fifty percent of the errors confusing velocity time graphs with position time graphs were caught by the modelers. However, the non-modelers did not catch any of these types of errors even when the actual number of errors, as in the case of the graphical error was greater in the non-modeling group. It seems highly unlikely that this would be a random event unrelated to the use of the modeling pedagogy. The increase in the number of errors caught might be one of the factors that allow the students to have greater problem solving success.

4.4.5.3 Metacognitive Changes Prior to Error, During Error Correction and After Error Correction

An exploration of the protocols was conducted to see if it could be determined if the participants who caught errors were doing something different between the time they made the error till they caught the error versus the times prior to and directly following this episode. The analysis below does not distinguish between modeling and non-modeling participants. Since only two non-modelers caught errors, it was decided to merge the analyze error corrections produced by all of the participants. One must be careful when looking at these results since only 41% of the modelers vs. 18% of the non-modelers caught errors. This means that only 32% of the entire group caught their physics errors. By comparing the verbal protocol data to the physics code the tasks completed from the time the error was made till when it was caught and the same number of statements directly before and after this episode could be analyzed. Using a 3 X 5 Chi-square table the types of statements made directly before, during and after the correction episode (time from making the error till the error is corrected) were tested for significance. Table 4-13 establishes a significant difference in the types of statements made during these time frames.

TABLE 4-13 Proportion of Statements made before the Physics Error Correction Episode was made vs. those made during the Physics Error Correction Episode vs. those made after the Physics Error Correction Episode

	Understand/read	Problem Analysis	Solving	Metacognitive	Miscellaneous	Total
Before Physics Error was made	13	24	6	12	2	57
Error Correction Episode	4	16	9	27	0	56
After Physics Error was made	12	23	7	14	1	57
Total	26	63	22	53	3	170

(Chi-square test = 14.73, $p < 0.01$, df = 4)

This demonstrates that the participants before the physics error is made are involved in problem analysis tasks 42% of the time. During this time frame understand/read and metacognition tasks each take up about 20% of the subject's time. When the error is made the task proportion changes and shifts towards 47% of all the tasks being metacognitive in nature while the problem analysis tasks drop to 28% and the understand/read tasks drop to 7%. The shift to debugging occurs very quickly and the protocol data did not seem to hint at what was triggering the change in student behavior. After the error is caught the task proportion shifts back to where it was before the error was made with 40% of the tasks being in the area of problem analysis. These transitions are illustrated in Figure 4- 5.

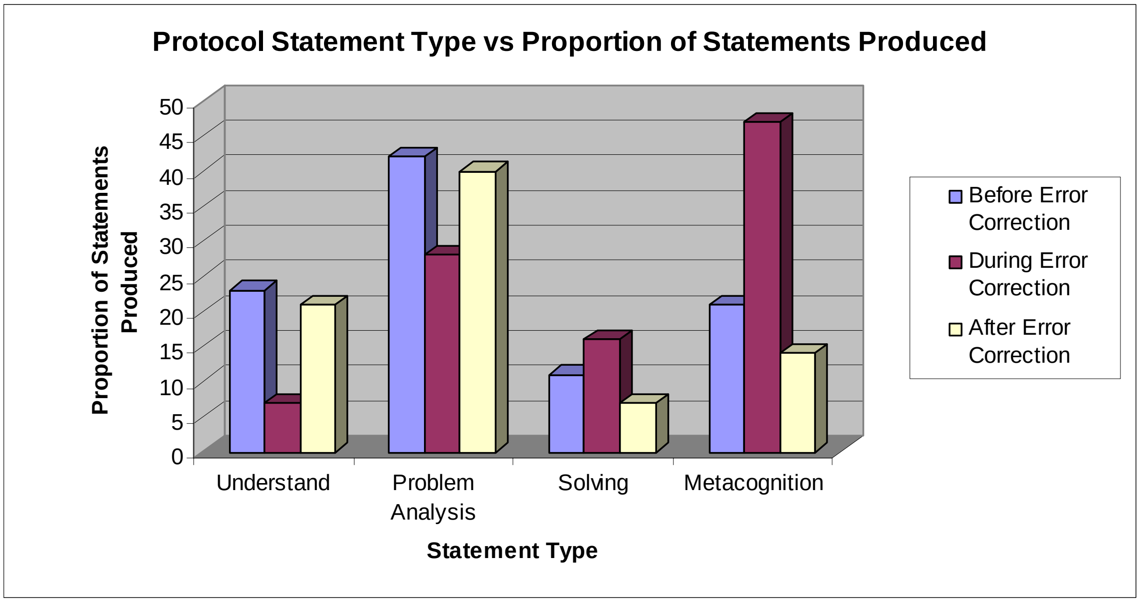


Figure 4-5: Statement Types vs. Proportion of Statements Produced

Further exploration reveals that during these time periods the type of metacognitive tasks the participants are involved in reveals a shift in attention. Prior to the error the majority of the metacognitive statements (50%) are in the area of monitoring comprehension. During and after the error correction episode the metacognitive tasks seem to be mostly in the area of evaluation and checking (about 60%). The major difference between the time during the error correction and the time after the correction is that a shift from 26% monitoring comprehension and 15% planning to 14% monitoring comprehension and 29% planning occurs while the percentage of evaluation remains basically constant. This is illustrated in Figure 4- 6.

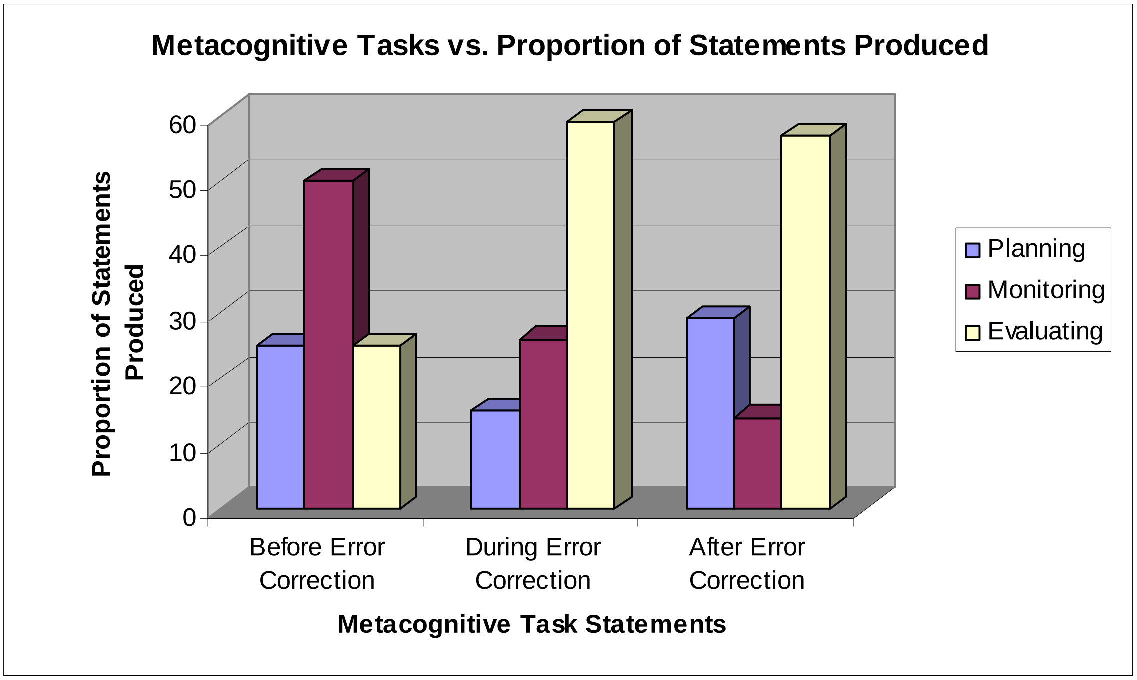


Figure 4-6: Metacognitive Statement Types vs. Proportion of Statements Produced

When one explores the difference in evaluating/checking statements one finds that prior to the error the majority of the evaluating is in the area of checking the information used and the equation chosen. After the error is made 75% of the participants' statements are in the realm is concentrated on evaluating the answer. Even after the correction is made the student still evaluates the answer to the same degree. However, after the error is corrected the percentage of evaluation of the approach increases from 6% to 25%. This information is summarized in Figure 4-7.

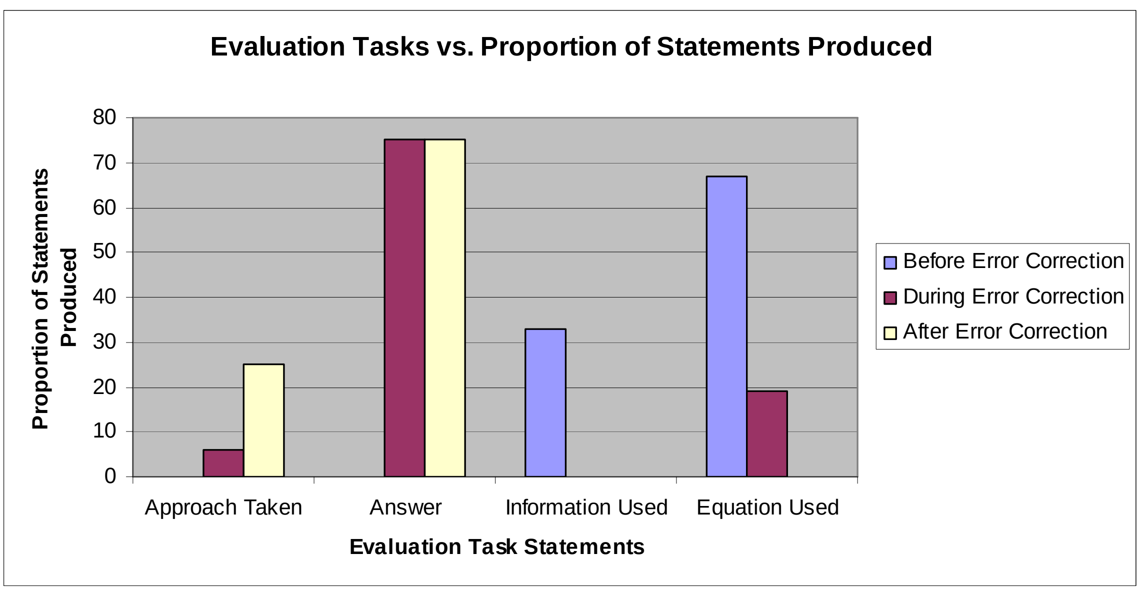


Figure 4-7: Evaluation Task vs. Proportion of Statements Produced

While evaluating the answer, the participants produce a similar proportion of statements (33%) during the error correction analyzing the answer by asking if it makes sense and by using alternate approaches. After the error is corrected the percentage of statements attempting to discover if the answer makes sense rises to 50% while they still check the answer via alternate approaches the same 33% of the time. It seems that the metacognitive skills used during and after the error aid in discovering if the error is substantive. In addition, after the error it seems that the students' continue to evaluate the answer checking if their decision about the error was indeed correct. The error correction changes should not be different between modeling and non-modeling students. However, one would expect that the metacognitive scaffolding that takes place in the course would allow for a greater number of modelers going through this process to correct their errors as was shown in the analysis.

4.4.6 Correlation Results between Study 1 and Study 2

It was predicted that the correlations between expert-like and novice-like behaviors observed in sections 4.4.1 – 4.4.5 would correlate with the students FCI, PS Task, expert and surface feature scores. In this chapter the first behavior that was analyzed was the time taken to complete the problems. Table 2-1 suggests that the time taken to complete problems would be less for expert-like students. As predicted the time taken to complete the five verbal protocol problems correlated quite well with the FCI ($r = -0.5$), expert ($r = -0.43$) and the PS Task score ($r = -0.54$) accounting for 25 %, 18 % and 29 % of the variance, respectively. All three of these values were negatively correlated meaning that as the scores increased on these measures the length of time to solve the problems decreased. The surface feature score was not significantly correlated to the time taken but was in the direction predicted.

The behaviors of identifying models vs. problem types, identifying algebraic representations vs. identifying equations, and working forward vs. working backwards are much more telling expert-like traits. In the initial problem analysis the subjects either chose to identify a problem type, identify a model or identify an equation to use to solve the problem. When identifying a model the students would key into the deep structure used in the problem and state whether the concept they were dealing with was momentum or energy for example. On the other hand, the non-modelers seemed to identify problem types and equations more often. The non-modelers would spend a majority of their time searching equation lists for specific formulas. When the non-modelers identified problem types they would key into surface features such as collisions or incline planes demonstrating a high reliance on surface features. They might see an object go down a slide and immediately looking for other incline plane problems to find a match. One would suspect that a high FCI and expert score would correlate positively with identifying a model and algebraic representations; whereas, a high surface feature score would correlate with identifying a problem type or equation. The correlations between the Study 1 task scores and problem analysis observed behaviors are summarized in Table 4-14. All of the regressions are in the direction predicted while only one is not significant: FCI score vs. identifying problem type. The FCI, expert and PS Task scores all were negatively correlated with identifying problem type, identifying an equation and working backwards towards a solution; whereas, the surface feature score was positively correlated with these behaviors. The reverse was true for identifying a model, identifying an algebraic equation, and forward chaining. All of these results are as predicted. A student's expert and surface feature scores account for a large proportion of the variation in these behaviors especially when identifying a model. It is clear that the students are using the knowledge structures they have constructed after the completion of the year long physics course to solve the problems in this study.

Table 4-14 Summary of Correlation Coefficient/R^2 values between post-tests and verbal protocol problem classification data.

Categories that are statistically significant at $p < 0.05$ are highlighted.

Verbal Protocol Problem Analysis Observed Behavior	Expert Score	FCI Score	Surface Feature Score	PS Task Score
Identify model	0.63/39.5	0.36/13.2	-0.51/26.4	0.45/20.5
Identify problem type	-0.38/14.4	-0.29/8.3	0.54/29.6	-0.45/20.1
Identify algebraic representation	0.47/21.7	0.56/31.4	-0.54/29	0.52/27.4
Identify equation	-0.36/12.8	-0.38/14	0.41/17.1	-0.58/34
Forward chaining	0.46/21	0.56/31.3	-0.44/19.5	059/34.4
Working backwards	-0.48/22.6	-0.64/41	0.69/47.3	-0.62/38.8

The data from study 2 concerning metacognitive behaviors did not show any pattern of significant correlation values with the task scores from Study 1. It is surprising that no correlation was found between the metacognitive and problem-solving behaviors since prior research has shown, as has this study, they are important skills to use in order to be a competent problem solver. This result demonstrates the need for more research in this area.

4.5 Discussion

4.5.1 Research question one

One of the main research questions for Study 2 was:

Will the "expert-like" knowledge structures of the modeling participants allow for more "expert-like" problem-solving strategies?

Overall, the findings of the verbal protocol study confirm that the modeling participants did have more "expert-like" problem-solving strategies which are in line with the more "expert-like" knowledge structures discovered in Study 1. There were twenty-two statistical tests run on the data to support this research question and my initial predictions. Fourteen of the tests were significant at the $p < 0.01$. Eight of the tests were significant with $p < 0.05$. Only five tests had p-values greater than 0.05 and four of those were in the direction of the initial predictions. Therefore, only one test was not in the direction predicted based on connection between expert-novice prior research and the modeling pedagogy. The results support the conclusion that modeling students are employing more expert-like strategies during problem-solving.

The first piece of evidence demonstrated that the modelers obtained more correct solutions than the non-modelers. This was similar to the finding in Study 1. Larkin et al. (1980a, 1980b) found that experts solved problems in half the time of novice participants. The evidence from this study showed that the modelers solved the problems more quickly than the non-modelers for six out of the seven problems completed (three of these being statistically significant). Finally, the analysis of the physics codes allows for the finding that the non-modelers produced significantly greater numbers of physics errors. These findings all point to a more "novice-like" non-modeling group and a more "expert-like" modeling group.

The conjecture that the modelers had more "expert-like" problem-solving strategies continued to be supported as the finer details of the protocol analysis were examined. The non-modelers problem-solving strategy proved to be quite novice in appearance as demonstrated in several areas. A large proportion of the non-modelers statements are referring to the problem statement (25%) with no progression towards a solution path. In addition, they were consistently referring to equation lists in order to find an equation in which to plug in their numbers. The data showed that the non-modelers in this study spend a high proportion (11%) of their problem analysis searching their notebooks or textbooks while the modelers only complete this task 2.2% of the time. In addition, the data showed that when solving a problem non-modelers identified equations 45% of the time when classifying a problem statement. Thus the non-modelers seem to be searching for either an equation in which to plug in their numbers or a sample problem to copy. During the retrospective interviews when discussing their general problem-solving strategy several of the members mentioned equation lists and their desire to plug in numbers in order to find the missing information. Some excerpts from the non-modelers are below:

NM22 said:

> Well I just kind of… Like it's always… Like I always have basically like my formulas – my formulas list is like key kind of. I just look for what I have, what I need and then try to plug it into a formula that's going to work.

NM11 seemed to confirm this sentiment when she said:

> What can I use all these pieces in that will get me to that final answer? Cause that seems to be the most physics problem that I've gotten are a bunch of pieces of little information and then I've got to find something that uses that information to get to my answer. So that's basically my strategy.

and finally NM4 continued by saying:

> Then I'll be like okay now I need to find an equation that has all them. And normally it's really easy because the problems that I have are the problems that we're learning and I know exactly where they are and I just use those. And that's how I solve them and then… the rest is just Algebra and I… I'm not that bad at it; sometimes I make mistakes – a lot of the time. Just keep them [referring to equations] from the beginning of the year to the end of the year and then maybe that will help me along the line somewhere where I have to …..

Several researchers (Chi et al., 1981, Hinsley et al., 1978 and Larkin et al., 1980) have shown that surface features are very supportive and reassuring for novice students. In this study, as in Study 1, it was also easy to see that the non-modelers were attending primarily to the surface feature of a problem. A large percent of the participants seemed to participate in a problem-matching strategy to solve the problems. This was seen quantitatively by the significant number of non-modelers who identified the problem type when classifying the problem to be solved. The non-modelers participated in this strategy 8% vs. 0.7% for the modelers. These percent proportions were calculated without using the retrospective interviews. When the retrospective interviews were included the percent proportion of non-modelers identifying problems by problem type doubles to 16%. This strategy was mentioned by NM18 in the retrospective interview:

I just… I guess I just look for clues, words or something that would remind me of some sort of problem that I did and I guess that's where I really start. , like a sample problem, yeah. That's… I do my problems like that. I think back to a sample problem really. That's why like… when the last test he gave us was an open… the last twenty minutes was an open notebook test. There was one or two problems that I just like… I knew that they were in my notebook, I knew kind of how to do them but I just like wasn't sure – I wasn't sure the order of what to do things in so I just said okay well I'll wait until we're allowed to use our notebooks and I'll just like change all the numbers basically – and that's what I did so… that's my personal strategy I guess.

Or as NM19 said:

try to remember past problems that I've done and how I went about like… sort of like patterns and things like that…. (*pause*)That's how my mind works…I think back to other problems we've done and I remember where it was and stuff and try to go from there. Sort of like a sample.

NM19 made a similar statement in the talk-aloud protocol when encountering a question that asks when a rabbit and a turtle would be in the same position. In the midst of reading the problem she said:

Oh I know we did some of these awhile ago dealing with trucks. (*flipping pages*) I am looking to find…(*flipping pages*) .. Okay, there it is

It seems she was very much seeking a problem that had similar literal surface features such as in this example a problem that had two objects meeting up with each other at some future date.

The analysis of the metacognitive task of evaluating the answer in terms of comparing it to a problem type when on a productive physics path demonstrated that the modeling participants look at deep structure while the non-modelers are swayed by surface features. This was supported when it was determined that none of the 19 modeling participants evaluated an answer in this fashion while three out of eleven of the non-modelers carried out this type of evaluation. This is another example of the greater "novice-like" problem-solving strategies of the non-modeling participants.

Another major type of novice problem-solving strategy that researchers have encountered is their tendency to work backwards towards a solution. The protocol study showed that the non-modelers worked backwards towards the solution 52% of the time. This type of strategy was seen in a number of protocols. Below is an excerpt from NM7's protocol when he was completing problem one and had identified a possible equation to utilize specifically; $v_f = v_i + a\Delta t$.

Okay I don't know my v_I… I know my acceleration; I don't know my time. Um… No not that one…[points to variables in equation $v_f = v_i + a\Delta t$ on board that he knows and does not know]... it's this one [points to an equation in his book. v_F squared equals v_I squared plus 2 a Delta x – there we go… [Writes $v_f^2 = v_i^2 + 2a\Delta x$]

Similar reflections can be observed in the retrospective interview which demonstrates the working backwards strategy such as when NM16 said:

I'd probably write down the information first. Then I'd have to think about what to do. Then I don't know. I always have my notebook. You know I always usually use the one [referring to equations] that has the variables that I want to solve for and not ones without them so… I mean if it has everything that I was given and what I'm looking for then that's the formula I use.

and NM7 said:

> Well I usually try and find a formula that would fit the problem and then… I try to limit my variables; if I can get it down to just one variable obviously that's the one I'm solving for so it's just easy math but, you know, if you can limit it to say like two equations that have a similar variable then you can redistribute and use algebra skills, or algebra

Given the more novice-like behavior above the non-modelers should have demonstrated that they were less flexible in their solution path. This was easily observed when it was found that the majority of the time (83%) they solved problems using algebraic solutions. This tendency to use algebraic solutions extended into the deployment of guess and check procedures which were always algebraic in nature.

The modeling group demonstrated problem-solving strategies that were the direct antithesis of the above. They were much more "expert-like" in all areas. Modelers were shown in this study to utilize graphical representations to a significantly greater extent than non-modelers. This is similar to past research that showed that experts utilized physical representations (i.e., diagrams) to a greater extent than novices (Larkin, 1977; Simon and Simon, 1978). The most telling quantitative finding in this study that demonstrates that the modeling participants are more "expert-like" is the finding that they identify the deep structure of the problems initially. The modelers verbally identified the models on which the problems were based 33% of the time. When the retrospective data is included this percentage increases to close to 50%. This finding is very much in line with prior expert-novice research showing that the experts are less swayed by surface features. In addition, this problem solving behavior lends support to the card-sort findings in Study 1 of this thesis. The retrospective statements of the modelers continue to support these findings. For example, M48 said:

> Well I mean I just… I try to think of all the models that we have learned. Well then with each model, you know, there's different equations. And I try to think of which equation all the information will fit into to give me one variable that is what I'm trying to solve for.

M24 stated the same sentiment but in a much more concise manner when she said:

> I guess I try to see what models I need to use.

Since the modelers are solving problems aided by a knowledge structure that is based on deep structure (i.e., models) then they have a smaller search space. They seem to complete a breadth search as M48 implies by looking at all six of their developed models to choose which one to utilize. Once the model is chosen they have available to them chunked with that model several useful representations that they can use such as algebraic representations and graphical representations. On the other hand the non-modelers have a much larger search space since they must rumble through equation after equation to find an appropriate one. This type of depth search is very inefficient.

Since the modelers identify the deep structure that the problem is based upon they should have a tendency to forward chain towards a solution. This was found to be true in this study when it was shown that the modelers forward chain 70% of the time which is significantly greater than the non-modelers. The retrospective again supports these findings. When M3 was working on Problem 1 he decided that he needed to use the constant acceleration model. In the retrospective M3 further explains that:

> Like, I didn't know the final velocity but after awhile it occurred to me that I could calculate the final velocities because the acceleration was… The acceleration due to gravity and we knew the time so we could calculate the final velocity. And then we could use that [pointing to the final velocity] into this algebraic model and solve for displacement.

This statement demonstrates how after identifying the model, members of the modeling group have at their disposal a number of algebraic representations to use to solve for the variable. They need not attempt to find one equation that has the unknown since they can move amongst the algebraic representations that are chunked with their model.

The identification of the deep structure of the problem via models should also allow the participants to have more flexibility in solution paths since the model should be chunked in their knowledge structures and include not only several algebraic solution paths but graphical solution paths as well. This was observed quantitatively when the modelers chose to solve problems graphically, 28% vs.7% for the non-modelers. On a number of occasions the modeler's use of graphical solutions was in itself an example of forward chaining. Take this except from M14's solution of problem 1 as he is drawing a velocity time graph of the situation:

> The acceleration is constant and then the area under the velocity curve tells you distance and so if It tells you displacement so if it's 9.8 slope… It doesn't… It doesn't… 50 kilograms doesn't matter (*inaudible*) 25 that has nothing to do with it. So it's 9.8 slope and it is for… Time is here so… and that goes to 9.5 seconds. And… for every second that it… its velocity it goes 9.8 more. So 9.8 times 9.5 is 93.1. So that's how high that is. [referring to highest velocity value on the y-axis on the graph]

At this point M14 goes on to calculate the area under the curve of the velocity time graph he constructed thus calculating the height of the building. Finally another piece of evidence concerning the flexibility of approach is directly observed when one notes that when the modelers choose to conduct guess and checks they always did so graphically while the non-modelers never graphically guess and check.

At almost every turn one observes more expert-like behavior on the part of the modeling participants. When quantitative data is brought to bear on this finding one finds that the significance level is usually at the $p < 0.01$ level. Therefore, there is substantial evidence to support the conjecture that the modeling pedagogy produced students with more "expert-like" knowledge structures than the non-modeling pedagogy which in turn produces more "novice-like" problem-solving strategies.

4.5.2 Research question 2

The second research question deals with the development of metacognitive skills on the part of the modeling participants and asked:

> What metacognitive differences exist between the two groups due to the modeling pedagogy?

The conjecture is that the modeling pedagogy should allow for the development of students who are more analytical thinkers, evaluating their approaches and conclusions. In order to answer this question Study 2 was designed to allow for the analysis of metacognition while participants were on a productive physics path. To analyze this research question twelve quantitative tests for significance were completed. Nine of these tests were significant. Three of these tests were not significant but were in the direction of the initial prediction. Therefore, all of the analyses completed were in the direction of the initial predictions.

It was determined that the modelers did indeed engage in a significantly greater proportion of metacognitive behaviors when on this productive path ($p < 0.03$). It was hypothesized that if one looked at the types of metacognitive behaviors' groups' exhibit one would find that the modelers would do more evaluative tasks. This was hypothesized because of the modeling pedagogy's implicit stress on analytical thinking (in this pedagogy students are continually asked to evaluate why they take specific solution paths and to evaluate if the findings are sensible). It was thought, since modeling pedagogy allows for the chunking of the model, that once the students decided on a specific model, they would not feel the need to monitor their comprehension very often. On the other hand, it is believed that the non-modelers, being uncertain of their path, would monitor their comprehension to a greater extent. Both of these findings were determined to be significant as shown in section 4.4.4. The modelers do evaluate their work across the entire solution path to a greater extent than the non-modelers.

Even through the modelers make more evaluating statements than the non-modelers the data allowed for an analysis of the differences in the type of evaluations made between the two groups. The first step in a problem-solving sequence would be deciding on the approach to take. It was determined that proportionally (i.e., fraction of statements) the modelers evaluated the approach at the same level as the non-modelers. Therefore, the question became what was different about how the two groups evaluated their approach. The only significant result as predicted, at $p < 0.01$, was that the modelers evaluated whether the approach made sense to them while the non-modelers would either review the steps in the approach or revise the approach. M16 demonstrated this sense making after she deployed an incorrect model for problem two when she said:

> It seems like if I don't know the height from which it was dropped I can't know the velocity and its momentum when it hit the water.

Since the use of the momentum model did not make sense, M16 revised her model selection and ultimately deployed the correct model: Newton's second law.

The next step in the solution path after evaluating the approach chosen by the subject would entail evaluating the final answer. The prediction that modelers would check the answer to a greater proportion was significant at $p < 0.05$. Due to the increased "expert-likeness" of the modeling participants' problem-solving strategies it was predicted that they would be more likely than the non-modelers to evaluate their answer via an alternate approach since the models developed would contain multiple solution methods. This was discovered to be a significant finding as the non-modelers choose not to evaluate their answers using this method. There are numerous examples of the modelers completing this evaluation such as the following excerpt from M14 after having solved problem 4C graphically:

> So then at 6 seconds for the turtle it is also 30 meters so that's when they catch up to each other but I'm not sure. I think there's a way to do it with algebra because it's 5 meters per second times x seconds ….meters and then ½ x seconds times the…
> [starts to write equation and writes 5 m/s*x_s = 1/2x_s*]

M14 then proceeds to check her answer via an algebraic approach. It was predicted that the analytical skills of the modelers would allow them to undertake a sense making evaluation of their final answer. While they did this 34% of the time it was not significantly different than the non-modelers. However, when one considered the number of participants undergoing this sense-making activity in both groups it was determined that a significant majority of modelers (68%)

were completing this task while a surprising small minority of non-modelers (27%) attempted to make sense of their answers. When looking at the modeling protocols it is easy to find examples of this type of evaluation of the answer such as when after completing problem one M1 says:

> So I multiply the two on this calculator and 9.5 times 95 is 902.5 meters. And now I'm going to think if that'll make sense.

It obviously did not make sense to M1 since she then said:

> So I solve for change in velocity. And I use the time and then once I have that I solve for Delta X.

M1 was checking the steps in her initial approach and did not find them lacking but the answer still did not make sense to her. At this point M1 then says:

> Now I'm going to see if any of the other algebraic models work.

At this point M1 then finds a different algebraic representation to utilize which allows her to obtain a height that makes sense to her. This type of analytical reasoning allowed a number of modeling participants to arrive at correct solutions.

This analysis shows that there is a distinct difference between the metacognitive skills of the two groups. The modelers evaluate what they are doing to a greater extent than the non-modelers even when on a productive path. In addition, they seem to try to make sense of the process to a greater extent. All of these skills should lead to the modelers becoming more "expert-like" as they develop greater problem-solving skills and discover their misconceptions using their problem-solving skills.

4.5.3 Research question 3

Study 2 was designed so that the physics errors produced could be analyzed in order to study the third major research question:

> Would the modeling participants commit fewer physics errors while catching a large proportion of those physics errors?

As has already been mentioned in the discussion about "expert-like" problem-solving strategies the modelers did commit fewer physics errors. However, the modelers caught a whopping 23 % of their physics errors! The modeling pedagogy allowed the participants to correct several errors that seemed to be very difficult for the non-modelers. These successfully corrected errors included discovering that one was confusing a velocity vs. time graph with a position vs. time graph and that one was confusing final velocity with average velocity. The non-modelers were unable to catch any of these types of errors. This difference between the two groups might be due to the greater flexibility in solution paths the modelers demonstrate and the greater metacognitive skills such that when they reach an answer they do not think they are finished but instead they attempt to evaluate the solution obtained. One might say that the modelers are simply more conversant with the use of graphs therefore catching more of the graphical type of errors since the graphical representation is a major proportion of each of the models they construct. However, this was disproved since in Study 1 the majority of both groups were able to solve the graphically based problems in the PS Task and in Study 2 all of the participants solved the initial graphical included in the protocol. While the increased analytical skills of the modelers might allow for these error revisions there were some errors that both groups seemed unable to correct. These errors were ones dealing with using variables from two dimensions in the same equation and not including all the forces in a net force problem. One would hope that as the modelers' progress they would continue to develop greater "expertness" allowing them to catch these errors. The development could be facilitated greatly by the production of deployments that required the students to draw diagrams in order to solve the problem. For example, problems that required the students to diagram the two dimensional situation would allow them to be more conscious that variables in different dimensions cannot be directly related to one another.

An analysis of the metacognitive activity of all the participants who caught errors implies that an increase in the metacognitive activity during the physics error correction episode (the time from when the error is made till it is corrected) in comparison to the same activity before and after the episode may be producing the error revisions. This is implied by the fact that the proportion of metacognitive tasks vs. the other three major problem-solving tasks (i.e., problem analysis, etc.) is the same both before and after the physics error correction episode. A more detailed analysis shows that within the metacognitive tasks the evaluating task increases greatly during and after the correction episode in comparison to the time before the error. Before the error the only evaluations completed are in the area of information used and equation appropriateness. This might be understandable since at this time most of the focus would be on setting up the approach. The increase in answer and approach evaluation after the episode might be due to the heightened awareness of the subject and their desire to make sure the correction was needed. Hence during both time periods, the participants were attempting to make sense of the answer and check it via alternate approaches. This analysis would seem to imply that when the level of evaluation is increased one would be able to correct a greater number of errors.

4.5.4 Research question 4

The fourth research question deals with how Study 1 and Study 2 are related to each other:

How do the results of Study 2 correlate with the findings of Study 1?

It was discovered that the tasks used in Study 1 can be utilized to predict a number of the problem-solving behaviors that student's exhibit. For example, it was discovered that the expert score can predict to some extent whether a student identifies a model to frame the solution and whether they work forward towards that solution. In addition, a high surface feature score allows one to predict that a student would identify problem types more often and work backwards towards a final solution. The task scores from Study 1 did not predict the student use of metacognitive skills. Clearly, more research in the area of metacognition and paper and pencil tasks that might be predictive of level of usage is needed.

4.5.5 In the final analysis

Study 2 proves that the modeling pedagogy produced students who perform more "expert-like" than the non-modeling participants who were trained in a traditional pedagogy. The expert-like problem-solving strategies of the modeling students allow for a greater number of correct answers possibly due to the production of fewer physics errors on average and the correction of a greater number of errors produced. All in all, the study shows that the modeling pedagogy is one method by which students can be guided along the path to becoming expert. The results of Study 2 demonstrated that Modeling Instruction in Physics is more effective in producing "expert-like" students than other more conventional instruction methods.

Chapter 5

Discussion and Conclusion

The main purpose of this thesis was to determine how the knowledge structures, metacognitive skills, and problem-solving abilities compared between modeling and non-modeling students. This chapter will review the results of the two studies designed to answer this basic question as well as discuss the theoretical and instructional implications.

5.1 Overview of Studies and their Results

Even though the Modeling Instruction pedagogy has been well documented to produce large conceptual and problem-solving gains in exiting students there had been no research into the changes in the knowledge structures developed by these students. Therefore, Study 1 was designed to determine if changes in the knowledge structure did occur and if these changes were correlated with increases in FCI and problem-solving task scores. The students for this study attended two modeling schools (one private and one public) and two non-modeling schools (one private and one parochial) and were taking first year trigonometry based physics classes. In addition, two second year classes, one modeling and one non-modeling, were incorporated into the study in order to look at the increase in conceptual understanding, problem-solving skill and knowledge structure organization over a two year period.

The organization of the students' knowledge structures were assessed using a card sort task similar to that used by Chi et al. (1981). An analysis of the differences observed between pre and post card sorts given to the first year and second year modeling based high school classes demonstrated that the students' knowledge organization became more similar to an expert modeler's knowledge structure and was based upon basic physic models. The card sort scores demonstrated that their structure continued to shift from a novice organization based upon surface features and literal problem elements to a more "expert-like" organization as time progressed.

In order to compare the differences and similarities between the modeling and non-modeling students' knowledge structure a confusion matrix was used to obtain an "expert-like" score. This score allowed for ease in comparison of the organization developed since the scores obtained for an expert-like sort ranged from 0 (novice) to 100 (expert). The scores generated showed that a modeling student's final knowledge organization was much more "expert-like" than that of a non-modeling student's by the end of the school year for both years. Surprisingly, the non-modeling second year physics students' did not demonstrate any additional shift towards a more "expert-like" knowledge organization than that which occurred for a first year non-modeling student! Comparisons of FCI and problem-solving task scores generally followed the trends of past research demonstrating that the modeling students obtained not only a better conceptual understanding but also improved problem-solving performance. Utilizing the card sort scores it was shown that the post-test expert knowledge structure correlated highly with a student's post-test FCI and PS Task score. In contrast a reliance on a structure based upon surface features correlated negatively with both. In addition, it was shown via a stepwise regression that the expert card sort score and FCI score seem to be predicting the same variance on the PS Task score meaning that they are testing similar developments in the student.

Study 1 was successful in demonstrating the development of a more "expert-like" knowledge structure but did not allow for one to observe how it was used during a problem-solving episode nor a determination of the problem-solving and metacognitive differences between the two groups. Therefore, Study 2 used a verbal protocol design to allow for the collection of detailed data in these areas using volunteers from the first year trigonometry based classes used in Study 1. Study 2 was quite successful in demonstrating that the modeling students utilized a more "expert-like" problem-solving strategy. They consistently identified the model they would use to solve the problem via a qualitative analysis similar to the expert novice research reviewed in Chapter 2 (Larkin, 1979; McDermott and Larkin, 1978). After identifying the model they would consistently identify an algebraic representation or graphical representation with which to solve the problem demonstrating that these seemed to be linked or chunked with the model (Larkin and Reif, 1979). The modelers' use of other than algebraic solutions demonstrated their ability to be more flexible problem solvers similar to the findings about expert problem-solving abilities (Santos, 1995). In addition, the modeling students consistently used forward chaining to work towards the final solution (Larkin, 1981; Simon and Simon, 1978; Larkin, McDermott, Simon and Simon, 1980a, 1980b) and rarely referred back to textbook problem types (Dhillon, 1998) or the problem statement (Hegarty et al., 1995). On the other hand, the non-modelers consistently demonstrated novice strategies as predicted by their knowledge structures as determined in Study 1. A regression analysis of the students' verbal protocol behavior and the scores they obtained on the tasks in Study 1 demonstrated a significant correlation between these tasks and the behaviors observed. For example, the surface feature score accounted for 41% of the variance when predicting if a student would work backwards towards the solution. On the other hand, the expert score accounted for 40% of the variance when predicting if a student would work forward towards the problem solution.

Study 2 also allowed for a detailed analysis of the metacognitive activities of the students especially when they were on a productive solution path. It was determined that the modelers routinely made a greater proportion of metacognitive statements, especially evaluating/checking statements, when on productive paths demonstrating the expert trait of consistently checking their solution paths (Dhillon, 1998; Schoenfeld, 1985). The modeling students evaluated the answer and approach to a greater extent and consistently checked to see if their answer made sense (13 out of 19 modelers vs. only 3 out of 11 non-modelers).

The detail obtained from the verbal protocol allowed for an analysis of the physics errors. The analysis demonstrated that the modelers not only made fewer errors on average but they also caught and corrected a greater percentage than the non-modelers. This seems a bit strange since one might expect the non-modelers to catch more errors since they would have a greater number of opportunities to do so. A number of the errors produced were similar between the two groups. Therefore it would seem that students have extreme difficulty with these areas - ignoring forces and confusing final velocity with average velocity. Modeling students confused velocity time graphs with position time graphs and used variables from different dimensions in the same equation to a lesser extent than non-modeling students. The modeling students also caught (although not to a significantly greater amount) more errors confusing the use of average velocity with final velocity and velocity time graphs with position graphs.

Finally, the protocol data allowed the shifts in metacognitive tasks prior to, during and after the error correction to be analyzed after collapsing all of the modeling and non-modeling students' data. It seems that prior to correcting an error solvers spend the majority of their time completing problem analysis tasks. When the error is made the students' shift towards a greater proportion of metacognitive tasks as if they intuitively feel uncomfortable with the path they are taking. After the error is corrected they shift back to the same task proportions observed prior to committing the error. In addition, during the error correction and directly after the error correction episode metacognitive statements lean heavily towards evaluation/checking statements; specifically, checking the answer.

5.2 Implications for Research Literature

This study was the first to show that the knowledge organization developed by modeling high school students was different from that of conventional students. Most of the studies concerned with modeling based curriculum looked only at problem-solving performance, conceptual understanding, understanding of models and modeling, and scientific reasoning. In addition, this study is the first to determine the problem-solving and metacognitive skills developed and used by students taught using a modeling based approach to science during problem-solving situations. The results are an important contribution to our understanding of the possible mechanisms modeling students use in order to produce the conceptual and problem performance gains observed. By understanding what they are doing, we as researchers can now tweak the methodology to further improve the trajectory of learning for our students.

In addition to the implications mentioned above, this study is one of the first to quantify the Chi et al. (1981) card sort task. The quantification of this task allows for the determination of expert-like, surface feature and question asked scores. These scores allow one to easily determine where a student is located on the path to "expertness". The use of these scores will allow future research in other domains to more easily correlate knowledge structure development to the abilities being tested.

Study 2's protocol analysis allowed for the quantification of the types and quantity of errors produced by the two different groups of students. This understanding should allow teachers and modeling researchers to develop curriculum based interventions that would help to allow students to more easily confront these areas. In addition, the study allowed for an analysis of the metacognitive details that occurred before, during and after an error correction episode. There has been no other study that has looked at error correction and its convergence with metacognition in this fashion. A study similar in some respects to this one was conducted by Allwood (1984) when he studied the error detection processes in statistical problem solving. In this study Allwood (1984) analyzed the talk aloud solutions given by 16 students solving 2 statistical problems. His analysis looked at the types of evaluative episodes the subjects undertook and found that the closer the error suspicion episode followed after the error the more likely that the error was detected. However, Allwood (1984) did not analyze his protocols in order to determine what specific metacognitive and problem-analysis activities were occurring during the error correction episode. This section of the thesis points towards the need for additional research in the area.

5.3 Implications for Instruction

Although this study demonstrated that the modeling students have a more "expert-like" knowledge structure and more "expert-like" problem-solving and metacognitive skills it also demonstrated that there is room for improvement. The protocol study definitely pointed out that the modeling students are on a trajectory from novice towards expertness and still demonstrate a number of novice qualities. The modeling instruction pedagogy needs to stress and model to a greater extent a number of expert-like traits such as the use of multiple representations, the identifying of a model or principle to help analysis problem statements, and the need to work forward towards a solution. The most distressing finding of the protocol analysis was the lack of usage of force diagrams amongst both groups. The students did not see the need to draw force diagrams and as one student said "they can easily be drawn in your head" implying that there was no need to draw them on paper. This is a major failing on the part of the modeling curriculum since if experts are flexible then other representations such as force diagrams are one of the reasons that they are better problem solvers. The modeling curriculum needs the addition of problem sets that require the students to draw force diagrams in order to succeed so that they can internalize the need for these types of representations.

This study demonstrates that the teaching of metacognitive skills is extremely important if students are to catch and correct errors. The final step of analyzing the solution must be modeled and coached to allow more students to internalize what to look for when evaluating an answer. If the use of metacognitive skills that good problem solvers demonstrated in this study were highlighted for students and practiced by them then their ability to detect and correct errors should increase.

5.3 Additional Questions and Future Research

This thesis has uncovered the following areas that require further research:

- Work needs to be completed looking specifically at how students use metacognitive skills to detect and correct errors in order to further the understanding of these particular processes. Based upon the results of a more in depth analysis in this area one can then test instructional interventions that might heighten the use of the strategies uncovered.
- The use of the card sort task would allow one to complete a longitudinal study looking at the trajectory of the change in knowledge organization over time. Is the change linear over the course of one school year or is there more of a quadratic relationship with a sudden rearranging of the organization? In addition, does the knowledge organization continue to become more expert-like even without additional classroom physics training? As the students move from a first to a second year physics course, how do their knowledge organizations continue to progress? The final question in this area concerns whether the modeling students will collapse the basic models over time in order to produce a true expert like structure based on fundamental principles. There was some evidence in the card sorts that suggest that this might be occurring since the better problem solvers had started to collapse models together such as constant velocity and constant acceleration. Would this collapse continue as students progress towards becoming physics experts and at what rate would it occur?
- The modeling students experienced the greatest difficulty with the sorting of problems dealing with energy. The reason behind this difficulty especially for modelers should be investigated since energy is a principle which threads though all science subjects and is an important model for students to master in order to understand the connections between the different science disciplines. Does this difficulty and the resulting knowledge

organization shift when the modelers are exposed to a modeling course with an energy theme?

REFERENCES

Aleven, V., and Koedinger, K. R., (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. Cognitive Science, 26, 147-179.

Allwood, C. M. (1984). Error detection processes in statistical problem solving. Cognitive Science, 8, 413-437.

Anderson, J., Greeno, J., Kline, P., and Neves, D. (1981). Acquisition of problem solving skill. In J.R. Anderson (ed.), Cognitive skills and their acquisition. Hillsdale, JH: Erlbaum.

Andersson, B. and Karrqvist, C. (1983). How Swedish pupils, aged 12 -15 years, understand light and its properties. European Journal of Science Education, 5 (4), 387-402.

August, D., Flavell, J., and Clift, R. (1984). Comparison of comprehension monitoring of skilled and less skilled readers. Reading Research Quarterly, 20(1), 39-53

Bagno, E. and Eylon, B. (1997). From problem solving to a knowledge structure: An example from the domain of electromagnetism. American Journal of Physics, 65 (8), 726- 736.

Bagno, E., Eylon, B. and Ganiel, U. (2000). From fragmented knowledge to a knowledge structure: Linking the domains of mechanics and electromagnetism. American Journal of Physics, 68 (7), S16 – S26.

Bielaczyc, K., Pirolli, P.L., and Brown, A (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. Cognition and Instruction, 13(2), 221-252.

Brewe, E. (2002). Inclusion of the energy thread in the introductory physics curriculum: An example of long-term conceptual and thematic coherence. Unpublished doctoral dissertation, Arizona State University, Phoenix.

Brown, A. (1978). Knowing when, where and how to remember: A problem of metacognition. In R Glaser (ed.), Advances in instructional psychology (Vol.1, pp. 77-165). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A., Campione, J. and Barclay, C. (1978). Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. Child Development, 50 (2), 501-512.

Caramazza, A., McCloskey, M. and Green, B. (1981). Naïve beliefs in “sophisticated” subjects: Misconceptions about trajectories of objects. Cognition, 9, 117-123.

Chabay, R. and Sherwood, B. (2006). Restructuring the introductory electricity and magnetism course. American Journal of Physics, 74 (4), 329 – 336.

Chabay, R. and Sherwood, B. (2004). Modern Mechanics. American Journal of Physics, 72 (4), 439 – 445.

Chabay, R. and Sherwood, B. (2002). Matters and Interactions I: Modern Mechanics and Matter and Interactions II: Electric and Magnetic Interactions. New York, NY: Wiley.

Champagne, A. B., Klopfer, L. E. and Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. American Journal of Physics, 48 (12), 1074 – 1079.

Champagne, A. B., Klopfer, L.E., Desena, A.T. and Squires, D.A. (1981). Structural representations of students' knowledge before and after science instruction. Journal of Research in Science Teaching, 18(2), 97-111.

Charness, N. (1979). Components of skill in bridge. Canadian Journal of Psychology, 33, 1-6.

Chase, W.G. (1982). Spatial representations of taxi drivers. In D. R. Rogers and J. A. Sloboda, editors, Acquisition of Symbolic Skills. Plenum Press, New York, 1982. pp. 391–405

Chase, W. G. and Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4 (1), 55-81.

Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (Ed.). Advances in instructional Psychology (pp. 161-238). Mahwah, N.J.: Lawrence Erlbaum Associates.

Chi, M.T.H, Bassok, M, Lewis, M, Riemann, P and Glaser, R (1989). Self-Explanations: How Students Study and use Examples in Learning to Solve Problems. Cognitive Science, 13, 145-182.

Chi, M. T. H., Feltovich, P., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chi, M.T. H., Glaser, R., and Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (ed.), Advances in the psychology of human intelligence, (Vol. 1). Hillsdale, N.J.: Erlbaum Associates.

Chi, M. T. H., & Koeske, R. (1983). Network representation of a child's dinosaur knowledge. Developmental Psychology, 19, 29-39.

Cheng, K. K., Thacker, B.A., and Cardenas, R. L. (2004). Using an online homework system enhances students' concepts in an introductory physics course. American Journal of Physics, 72 (11), 1447-1453.

Clement, J. (1982). Students' preconceptions in elementary mechanics. American Journal of Physics, 50, 66-71.

Clement, J. (1991). Nonformal reasoning in experts and in science students: the use of analogies, extreme cases and physical intuition. In J.F. Voss, D. N. Perkins and J.W. Segal (eds.), Informal Reasoning and Education (pp. 345-362). Hillsdale, N.J.: Lawrence Erlbaum.

Cohen, R., Eylon, B. and Ganiel, U. (1983). Potential difference and current in simple circuits: a study of students' concepts. American Journal of Physics, 51, 407-412.

Collins, A., Brown, J. and Newman, S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (ed.) Knowing, learning and instruction: Essays in honor of Robert Glaser. Hillsdale, NJ: Erlbaum.

Cummings, K., Marx, J., Thornton, R. and Kuhl, D. (1999). Evaluating innovation in studio physics. American Journal of Physics, 67 (7), S38 -S44.

de Groot, A. D. (1965). Thought and choice chess. The Hague: Mouton.

de Jong, T., and Ferguson-Hessler, M. G. M. (1986). Cognitive structures of good and poor problem solvers in physics. Journal of Educational Psychology, 78(4), 279-288.

Desbien, D. (1987). Modeling discourse management compared to other classroom management styles in university physics. Unpublished doctoral dissertation, Arizona State University, Phoenix.

Dhillon, A. (1998). Individual differences within Problem-solving Strategies Used in Physics. Science Education, 82 (3), 379-405.

Ding, L., Chabay, R., Sherwood, B. and Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. Physical Review Special Topics – Physics Education Research, 2, 010105.

DiSessa, A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. Cognitive Science, 6, 37-75.

Dufresne, R. J., Gerace, W. J., Hardiman, P. T. and Mestre, J. P. (1992). Constraining Novices to Perform Expertlike Problem Analyses: Effects on Schema Acquisition. The Journal of the Learning Sciences, 2(3), 307-331.

Engelhardt, P., and Beichner, R. (2004). Students' understanding of direct current resistive electrical circuits. American Journal of Physics Teachers, 72 (1), 98-115.

Eylon, B. and Reif, F. (1984). Effects of knowledge organization on task performance. Cognition and Instruction, 1, 5-44.

Ferguson-Hessler, M. G. M. and de Jong, T. (1987). On the quality of knowledge in the field of electricity and magnetism. American Journal of Physics, 55 (6), 492-497.

Ferguson-Hessler, M. G. M. and de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. Cognition and Instruction, 7, 41-54.

Finegold, M. and Mass, R. (1985). Differences in the Processes of Solving Physics Problems between Good Physics Problem Solvers and Poor Physics Problem Solvers. Research in Science and Technological education, 3, 59-67.

Flavell, J. (1976). Metacognitive aspects of problem solving. In Resnick, L. (Ed.) The Nature of Intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(1), 906-911.

Galili, I., Bendall, S. and Goldberg, F. (1993). The effects of prior knowledge and instruction on understanding image formation. Journal of Research in Science Teaching, 30 (3), 271-301.

Gobbo, C. and Chi, M. T. H. (1986). How knowledge is structured and used by expert and novice children. Cognitive Development, 1, 221-237.

Gobbo, C., & Chi, M. T. H. (1986). How knowledge is structured and used by expert and novice children. Cognitive Development, 1, 221-237.

Goldberg, F. M. and Anderson, J. H. (1989). Student Difficulties with graphical representations of negative values of velocity. The Physics Teacher, 27 (4), 254 – 260.

Goldberg, F. M. and Bendall, S. (1995). Making the invisible visible: A teaching/learning environment that builds on a new view of the physics learner. American Journal of Physics, 63, 978-991.

Goldberg, F. M. and McDermott, L.C. (1986). Student difficulties in understanding image formation by a plane mirror. The Physics Teacher, 24, 472-480.

Goldberg, F. M. and McDermott, L.C. (1987). An investigation of student understanding of the real image formed by a converging lens or concave mirror. American Journal of Physics, 55 (2), 108-119.

Gunstone, R. F. (1987). Student understanding in mechanics: A large population survey. American Journal of Physics, 55, 691-696.

Gunstone, R. F. and White, R. T. (1981). Understanding of gravity. Science Education, 65 (3), 291 – 299.

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

Halloun, I. and Hestenes, D. (1985a). The initial knowledge state of college students. American Journal of Physics, 53, 1043-1055.

Halloun, I. and Hestenes, D. (1985b). Common sense concepts about motion. American Journal of Physics, 53, 1056-1065.

Halloun, I. and Hestenes, D. (1987). Modeling instruction in mechanics. American Journal of Physics, 55 (5), 455 - 462.

Hardiman, P.T., Durfresne, R. and Mestre, J. P. (1989). The relation between problem categorization and problem solving among novices and experts. Memory and Cognition, 17, 627-638.

Hayes, J.R. (1981). The complete problem solver. Philadelphia: Franklin Institute: Franklin Institute Press.

Hegarty, M., Mayer, R.E. and Monk, C.A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. Journal of Educational Psychology, 87, 18-32.

Heller, J. I., and Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. Cognition and Instruction, 1(2), 177-216.

Heller, P. and Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. American Journal of Physics, 60 (7), 637- 644

Heller, P., Keith, R., and Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. American Journal of Physics, 60, 627-636.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. American Journal of Physics, 55 (5), 440-454.

Hestenes, D. (1992). Modeling games in the Newtonian world. American Journal of Physics, 60 (8), 732 – 748.

Hestenes, D. (1996). Modeling Method for Physics Teachers. Proceedings of the International Conference on Undergraduate Physics, 935-958.

Hestenes, D., & Wells, M. (1992). A mechanics baseline test. The Physics Teacher, 30, 159-166.

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

Hinsley, D. H., Hayes, J. R., and Simon, H. A. (1977). From words to equations – meaning and representation in algebra word problems. In M. Just and P. Carpenter (eds.), Cognitive process in comprehension (pp.89-106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hmelo-Silver, C. and Pfeffer, M. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors and functions. Cognitive Science, 28, 127-138.

Huffman, D. W. (1994). The effect of explicit problem solving instruction on students' conceptual understanding of Newton's laws. Unpublished doctoral dissertation, University of Minnesota, Twin Cities.

Huffman, D. W. (1997). Effect of explicit problem solving instruction on high school students' problem-solving performance and conceptual understanding of physics. Journal of Research in Science Teaching, 34 (6), 551-570.

Kalman, C. S., Rohar, S., and Wells, D. (2004). Enhancing conceptual change using argumentative essays. American Journal of Physics, 72 (5), 714-717.

Karplus, R. (1977). Science Teaching and the Development of Reasoning. Journal for Research in Science Teaching, 14, 169-175.

Keith, R.L. (1993). Correlation between the consistent use of a general problem solving strategy and the organization of physics knowledge. Unpublished doctoral dissertation, University of Minnesota.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. Educational Psychologist, 27(1), 111-126.

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. Discourse Processes, 16 (2), 193-2002.

Klahr, D. and Carver, S. M. (1988). Cognitive objectives in LOGO debugging curriculum: Instruction, learning and transfer. Cognitive Psychology, 20 (3), 362-404.

Krutetskii. V. A. The psychology of mathematical abilities in schoolchildren. Chicago: University of Chicago Press, 1976.

Larkin. H. (1979). Processing information for effective problem solving. Engineering Education, 1979, 70(3), 285-288

Larkin, J. (1980). Skilled Problem solving in Physics: A hierarchical planning model. Journal of Structural Learning, 6, 271-297.

Larkin, J. H. (1981). Enriching formal knowledge: a model for learning to solve textbook physics problems. In J. R. Anderson (ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum Associates, 311-334.

Larkin, J. H., McDermott, J., Simon, D. P., and Simon, H. A. (1980a). Models of competence in solving physics problems. Cognitive Science, 4, 317-345.

Larkin, J. H., McDermott, J., Simon, D. P., and Simon, H. A. (1980b). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Larkin, J. and Reif, F. (1979), Understanding and Teaching Problem-Solving in Physics. European Journal of Science Education, 1(2), 191-203.

Lawson, M. and Fueloep, S. (1980). Understanding the purpose of strategy training. The British Journal of Educational Psychology, 50, 175-180.

Lawson, R. A. and McDermott, L. C. (1987). Student understanding of the work-energy and impulse theorems. American Journal of Physics, 55 (9), 811 – 817.

Leonard, W.J., Dufresne, R.J., and Mestre, J.P. (1996). Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems. American Journal of Physics, 64(12), 1495 – 1503

Lewis, A.B. (1989). Training students to represent arithmetic word problems. Journal of Educational Psychology, 81 (4), 521-531.

Maloney, D. P. (1984). Rule-governed approaches to physics – Newton's third law. Physics Education, 19, 37-42.

Maloney, D. P. (1994). Research on problem solving: Physics. In D. L. Gabel (Ed.), Handbook of research on science teaching and learning (pp. 327-356). New York, NY: Macmillan.

McCloskey, M. (1983). Intuitive physics. Scientific American, 249(4), 122-130.

McCloskey, M., Caramazza, A., and Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. Science, 1980, 210(5), 1139-1141.

McDermott, J. and Larkin, J. H. (1978). Re-representing textbook physics problems. In Proceedings of the 2nd National Conference, the Canadian Society for Computational Studies of Intelligence. Toronto: University of Toronto Press, 1978. pp. 156-164

McDermott, L. and Shaffer, P. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of students understanding. American Journal of Physics, 60, (11), 994-1002.

McKeithen, Katherine B., Reitman, Judith S., Rueter, Henry H. and Hirtle, Stephen C. (1981). Knowledge Organization and Skill Differences in Computer Programmers. Cognitive Psychology, 13, 207-325.

McMillan, C., and Swadener, M. (1991). Novice use of qualitative versus quantitative problem solving in electrostatics. Journal of Research in Science Teaching, 28(8), 661-670.

Mestre, J., Dufresne, R.J., Gerace, W.J., Hardiman, P.T., and Touger, J.S. (1993). Promoting skilled problem solving behavior among beginning physics students. Journal of research in science teaching, 30, 303-317.

Mevarech, Z. (1999). Effects of metacognitive training embedded in cooperative settings on mathematical problem solving. The Journal of Educational Research, 92 (4), 195-205.

Mevarech, Z. and Kramarski, B. (1997). IMPROVE: a multidimensional method for teaching mathematics in heterogeneous classrooms. American educational research journal, 34, 365-394.

Morote, E. and Pritchard, D. E. (2002). What course elements correlate with improvement on tests in introductory Newtonian mechanics? Paper presented at the Annual National Association for Research Science Teaching meeting, New Orleans, Louisiana.

Moss, J, Kotovsky, K., and Cagan, J. (2006). The role of Functionality in the mental representations of engineering student: some differences in the early stags of expertise. Cognitive Science, 30, 65-93.

Nathan, M.J., Mertz, K., and Ryan, B. (1994). Learning through self-explanation of mathematical examples: Effects of cognitive load. Paper presented at the 1994 Annual meeting of the AERA.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice Hall.

Neto, A. and Valente, M.O. (1997). Problem Solving in Physics: Towards a Metacognitively Developed Approach. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, 70th, Oak Brook, Il, March 21-24.

Neuman, Y., Leibowitz, L., and Schwarz, B. (2000). Patterns of Verbal Mediation during problem solving: A Sequential Analysis of Self-Explanation. The journal of Experimental Education, 68(3), 197-213.

Palincsar, A. S. and Brown, A. L. (1984). Reciprocal Teaching of Comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1(2), 117-175.

Paris, S. and Winograd, P. (1990). How Metacognition Can Promote Academic Learning and Instruction. In Jones, B. and Idol, L. (eds.), Dimensions of Thinking and Cognitive Instruction (pp.15-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Piaget, J. (1963). The psychology of intelligence. New York: Routledge.

Piaget, J. (1970). Genetic epistemology. New York: W.W. Norton & Company.

Pirolli, P and Bielaczyc, K. (1989). Empirical Analyses of Self-Explanation and Transfer in Learning to Program. Proceedings of the 11th Annual Conference of the Cognitive Science Society (pp. 450-457).

Pirolli, P.L., and Recker, M. (1994). Learning strategies and transfer in the domain of programming. Cognition and Instruction, 12, 235 – 275.

Polya, G. (1968). How to solve it. Princeton, NJ: Princeton University Press.

Priest, A.G. and Lindsay, R.O. (1992). New Light on Novice-Expert differences in physics problem solving. British Journal of Psychology, 83, 389-405.

Raghavan, K., Sartoris, M., Schunn, C., and Scott, K. (2005). Middle-School Students' Perceptions and Interpretations of Different Model Types. Paper presented at the annual NARST meeting, Dallas, Texas.

Reif, F. and Allen, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. Cognition and Instruction, 9(1), 1-44.

Reif, F. and Larkin, J. (1979), Understanding and Teaching Problem-Solving in Physics. European Journal of Science Education, 1(2), 191-203.

Reif, F., Larkin, J. and Brackett, G. C. (1976). Teaching general learning and problem-solving skills. American Journal of Physics, 44 (3), 212 – 217.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. Cognitive Science, 21 (1), 1-29.

Robertson, William C. (1990). Detection of cognitive structure with protocol data: predicting performance on Physics transfer problems. Cognitive Science, 14, 253-280.

Sabella, M and Redish, J (in press). Knowledge organization and activation in physics problem-solving. Submitted for publication in the Physics Education Research Supplement to the American Journal of Physics.

Salomon, G., Globerson, T and Guterman, E. (1989). The computer as a zone of proximal development: Internalizing reading-related metacognitions from a reading partner. Journal of Educational Psychology, 81(4), 620-627

Santos, M. T. (1995). Students' recognition of structural features in mathematical problem solving instruction. Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.

Savelsbergh, E.R, de Jong, T., Ferguson-Hessler, M.G.M. (1996). Forms of problem representation in physics (Report No. IST-MEMO-91-01). Enschede, The Netherlands: University of Twente.

Schoenfeld, A. H.(1983). In Lesh, R. and Landau, M (eds.), Acquisition of Mathematics Concepts and Processes (pp. 345 - 395). Orlando, Fl: Academic Press.

Schoenfeld, Alan (1985). Mathematical Problem Solving. Orlando, Fl: Academic Press, Inc.

Schoenfeld, A. (1987). What's all the fuss about metacognition? In A. Schoenfeld (ed.), Cognitive science and mathematics education (pp. 189-215). Hillsdale, NJ: Lawrence Erlbaum.

Schoenfeld, Alan H. (1992). Learning to think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. Grouws (ed.), Handbook of Research on Mathematics Teaching and Learning (pp. 334 – 370). New York, New York: Macmillan.

Schoenfeld, A. H. and Herrmann, D.J.(1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982b, 8(5), 484-494.

Schwarz, C.V. and White, B. (2005). Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling. Cognition and Instruction, 23 (2), 165-205.

Shavelson, R.J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 63 (3), 225-234.

Shavelson, R.J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. Journal of Research in Science Teaching, 11 (3), 231 - 249.

Shavelson, R.J. and Stanton, G. C. (1975). Construct validation: Methodology and application to three measures of cognitive structure. Journal of Educational Measurement, 12 (2), 67-85.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. Cognitive Psychology, 28, 225-273.

Siegler, R.S. (2000). Microgenetic studies of self-explanation. In Granott, N. and Parziale, J (eds.), Microdevelopment: Transition Processes in Development and Learning (pp. 31-58). Cambridge: Cambridge Press.

Silver, E. A. (1979). Student perceptions of relatedness among mathematical verbal problems. Journal for research in Mathematics education, 10 (3), 195-210.

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. Journal for Research in Mathematics Education, 12 (1), 54-64.

Simon, D. P., and Simon. H. A. (1978). Individual differences in solving physics problems. In R. Siegler (ed.), Children's thinking: What develops? Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978. 325-348

Singh, C. (2002). ''When physical intuition fails,'' American Journal of Physics, 70, 1103–1109.

Smith, M. (1992). Expertise and organization of knowledge: unexpected differences among genetic counselors, faculty and students on problem categorization tasks, Journal of research in science teaching, 29(2), 179-205.

Smith, E.E. and Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. Cognition and Instruction, 1, 359-396.

Snyder, J.L. (2000). An investigation of the knowledge structures of experts, intermediates and novices in physics. International Journal of Science Education, 22(9), 979-992.

Swackhamer, G. and Dukerich, L. (1996). Models and Modeling in the High School Physics Classroom. *Proceedings of the 1996 American Association of Physics Teachers summer conference*, College Park, MD. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257-285.

Sweller, J. and Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction, 2, 59-89.

Tabachneck, H. J. M., Leonardo, A. M., and Simon, H. A. (1994). In A. Ram and K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society.* Hillsdale, N. J.: Lawrence Erlbaum Associates.

Thacker, B. A., Ganiel, U., and Boys, D. (1999). Macroscopic phenomena and microscopic processes: Student understanding of transients in direct current electric circuits. Physics Education Research, American Journal of Physics Supplement, 67 (7), S25 – S31.

Thro, M. P. (1978). Relationship between associative and content structure of physics concepts. Journal of Educational Psychology, 70 (6), 971-978

Trowbridge, D. E. and McDermott, L. C. (1980). Investigation of student understanding of the concept of velocity in one dimension. American Journal of Physics, 48, 1020-1028.

Trowbridge, D. E. and McDermott, L. C. (1980). Investigation of student understanding of the concept of acceleration in one dimension. American Journal of Physics, 49, 242 - 253.

Van Heuvelen, A. (1991). Overview, case study physics. American Journal of Physics, 59, 898-907.

VanLehn, K. and Jones, R.M. (1992). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? Proceedings of the 15th annual Conference of Cognitive Science Society

Veenman, M. V.J. and Verheij, J. (2003). Technical students' metacognitive skills: relating general vs. specific metacognitive skills to study success. Learning and Individual Differences, 13, 259-272.

Veldhuis, G. H. (1986). Differences in the categorization of physics problems by novices and experts. Unpublished doctoral dissertation, Iowa State University, Ames.

Vesenka, J., Beach, P., Munoz, G., Judd, F., and Key, R. (2002). A comparison between traditional and "modeling" approaches to undergraduate physics instruction at two universities with implications for improving physics teacher preparation. Journal of Physics Teacher Education Online, 1(1), 3-7.

Viennot, L. (1979). Spontaneous reasoning in elementary mechanics. European Journal of Science Education, 1, 205-221.

Vygotsky, Lev (1962). Thinking and Speaking. Cambridge: The M.I.T. Press.

Webb, N.M. (1989). Peer interaction and learning in small groups. International Journal of Education Research, 13, 21-39.

Weber, Keith (2001). Investigating and teaching the strategic knowledge needed to construct proofs. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Weiser, M. and Shertz, J. (1983). Programming problem representation in novice and expert programmers. International Journal of Man-Machine Studies, 19. 391-398.

Wells, M. (1987). Modeling instruction in high school physics. Unpublished doctoral dissertation, Arizona State University, Phoenix, AZ.

Wells, M., Hestenes, D. and Swackhamer, G. (1995). A modeling method for high school physics instruction. American Journal of Physics, 63 (7), 606-619.

White, B.Y. (1983). Sources of difficulty in understanding Newtonian mechanics. Cognitive Science, 7, 41-65.

White, B. (1993). ThinkerTools: Causal Models, Conceptual Change, and Science Education. Cognition and Instruction, 10(1), 1-100.

White, B., and Frederiksen, J. R. (1998). Inquiry, Modeling and Metacognition: Making Science Accessible to All Students. Cognition and Instruction, 16(1), 3-118.

White, B. and Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. Educational Psychologist, 40(4), 211-223.

Wright, D. S. and Williams, C. D. (1986). A WISE strategy for introductory Physics. The Physics Teacher, 24 (4), 211- 216.

Zajchowski, R. and Martin, J. (1993). Differences in the Problem Solving of Stronger and Weaker Novices in Physics: Knowledge, Strategies and Knowledge Structures? Journal of Research in Science Teaching, 30 (5), 459-470.

Zhang, B., Wu, H., Fretz, E., Krajcik, J., Marx, R., Davis, E. and Soloway, E. (2002). Comparison of modeling practices of experts and novice learners using a dynamic, learner-centered modeling tool. NARST 2002 Strand 7: Educational Technology.

APPENDIX A: Experimenter's Test Instructions for Study 1

I. Instructions for Force Concept Inventory:

1) Do not write on the questionnaire.
2) Mark your answers on the answer sheet
3) Do not skip any questions
4) Avoid guessing. Your answers should reflect what you personally think.
5) Plan to finish this questionnaire in 30 minutes

II. Instructions for Card Sort Task:

The subjects are not allowed to use a pencil or pen and should not actually solve the problems.

Read the problems carefully and think about how you would go about solving them but do not actually attempt a solution. Then sort the problems into groups or categories based upon the solution method you would use to solve the problem. There are no maximum or minimum categories (i.e., if you wish you could make a category with only one problem). You may take as much time as you like to sort the problems.

After the sorting is finished in writing on the sheet of paper given you by the instructor explain your reasons for each grouping. Attempt to give each group a simple title that could be a single word or phrase or a short sentence. When finished titling each group paperclip the group together with the title card on top. When finished titling all the groups place a rubber band around the entire deck of cards with your name on top.

III. Instructions for Problem Solving Task:

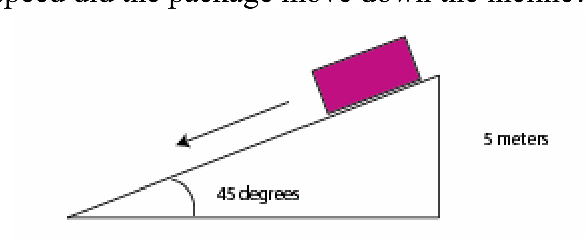
1) Answer the following questions to your best ability.
2) Show all of your work in the space provided.
3) Plan to finish the task in 40 minutes.

APPENDIX B: Card Sort problems delineated by Model Type

I. Constant Velocity Model:

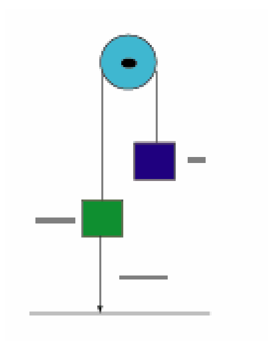
Problem 1.1
A person drops a feather off the top of a 20-meter tall building. Assuming that the feather reaches terminal velocity instantly, what is its velocity if it takes the feather 5 seconds to reach the ground?

Problem 1.2:
A 2.00 kg package is sliding down a 45-degree incline with a vertical drop of 5 meters. The package has an acceleration equal to zero. If it takes the package 2 seconds to reach the bottom of the incline with what speed did the package move down the incline?



Problem 1.3:
A compressed spring pushes a marble in a pinball machine. The spring gives the marble the ability to travel a distance of 1 cm every 0.1 seconds in a frictionless channel. If the channel is 1.25 meters long how long will it take the marble to reach the end of the channel?

Problem 1.4
A rope that passes over a massless frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. Due to a large amount of friction in the pulley the basket of bricks moves downwards at a velocity of 3 m/s. How far does the basket move in 5 seconds?

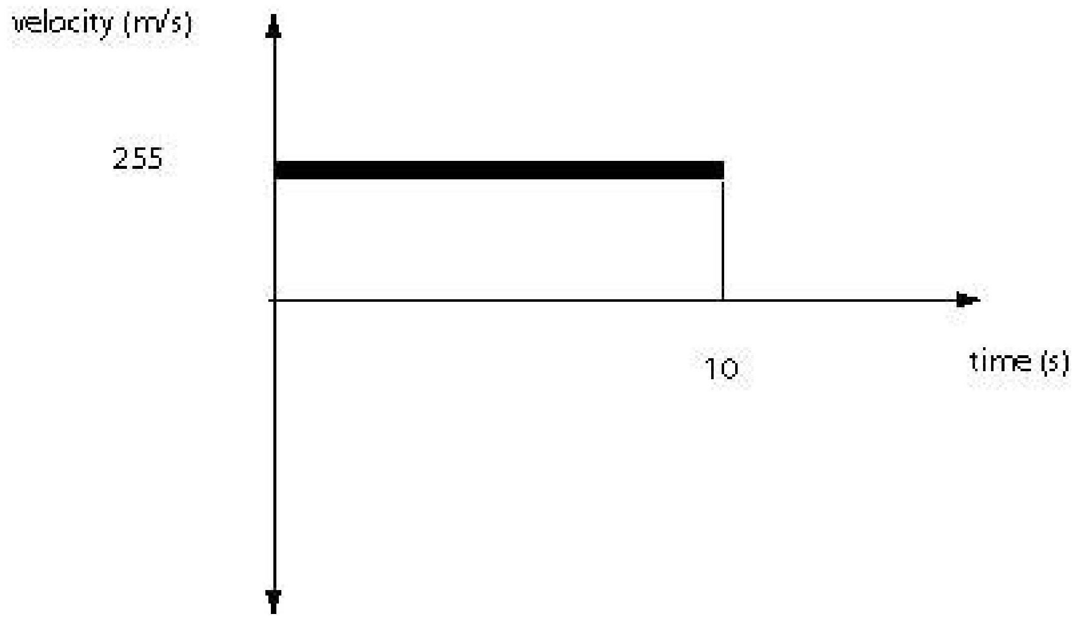


Problem 1.5
A batter in the World Series picks up a bat and carries it across the field at a rate of 5 m/s. If it takes the batter 3 seconds to reach home plate how far did the batter walk?

Problem 1.6
A 100 kg person is in an elevator moving upward at 2.5 m/s. How far does the person go in 10 seconds?

Problem 1.7
A car whose initial speed is 30 m/s continues to move down the road for 5 seconds. Determine the position of the car at the beginning and the end of the third second.

Problem 1.8
The motion of a parachutist is plotted on a velocity time graph below. Interpret the motion of the parachutist.

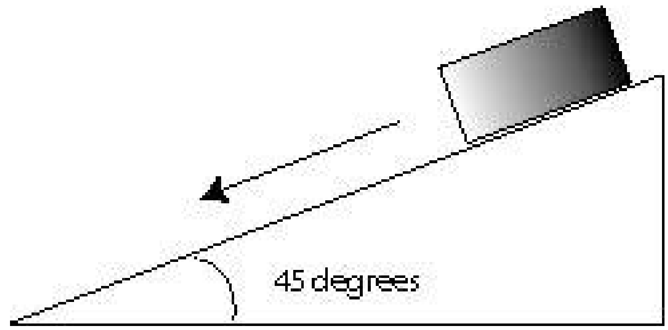


Problem 1.9
A bright physics student leaves his house at six a.m. to go to school. He travels directly there at 35 miles per hour. After school he returns directly home at a velocity of 40 miles per hour. What was his average velocity for the entire day?

II. Constant Acceleration Model:

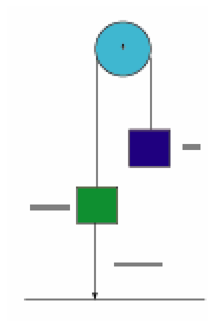
Problem 2.1:
Two people named Jim and Jane jump off a freeway overpass into a lake below. Jane has a mass of 120 kg and Jim is twice as massive as Jane. If it takes Jane 1 second to reach the lake how high was the overpass?

Problem 2.2:
A 2.00 kg package is released from rest at the top of a 45-degree frictionless incline. If it takes the package 1 second to cover a distance of 7 meters what was the packages acceleration?



Problem 2.3:
A spring inside a dart gun exerts a nonconstant force on a dart that gives the dart a velocity of 5 m/s as it leaves the muzzle of the gun. If the process takes 1.2 seconds what is the acceleration rate of the dart.

Problem 2.4:
A rope that passes over a massless frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest and experiences a net acceleration of 6 m/s/s. How far does the basket move in 3 seconds?

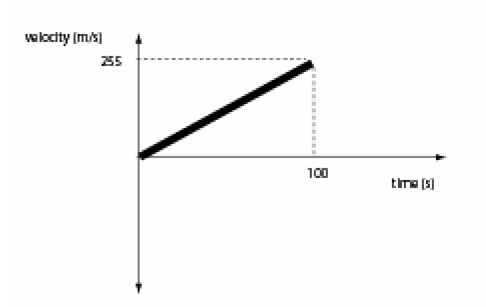


Problem 2.5:
A batter is struck out in the last inning of the World Series and looses the game. He throws his bat initially at rest with an acceleration rate of 10 m/s/s at the pitcher. If the batter is in contact with the bat for 8×10^{-2} seconds what is the final velocity of the bat as it leaves the batter's hand?

Problem 2.6:
A 100 kg person is in an elevator that started from rest and then moved upward at 2 m/s/s. How far does the elevator move in 10 seconds?

Problem 2.7:
A car whose initial speed is 30 m/s slows uniformly to 10 m/s in 5 seconds. Determine the position of the car at the beginning and the end of the third second.

Problem 2.8:
The motion of a parachutist was plotted on the graph below after she jumped from a plane. Describe the motion of the parachutist



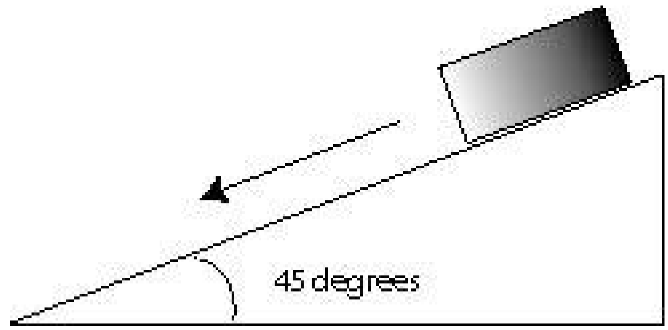
Problem 2.9
A paint ball gun supplies a constant force for 0.2 meters on a paintball inserted in its barrel. What is the acceleration rate of the paintball if it reaches a final velocity of 5 m/s at the end of the barrel?

Problem 2.10
A bus moving at 20 m/s (t = 0) slows at a rate of 4 m/s each second. How long does it take the bus to stop?

III. Model of Newton's second law:

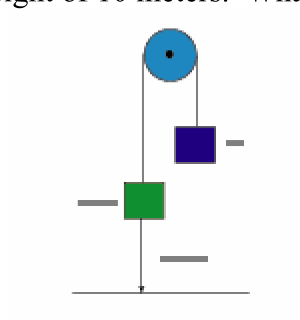
Problem 3.1:
Two people named Jim and Jane jump off a freeway overpass into a lake 10 meters below. Jane has a mass of 120 kg and Jim is twice as massive as Jane. If Jane slows down to a stop below the surface of the water at a rate of 2 meters per second per second what net force did the water exert on her?

Problem 3.2:
A 2.00 kg package is released at the top of a 45-degree incline that has a vertical drop of 5 meters. The coefficients of friction between the package and the incline are $\mu_s = 0.40$ and $\mu_k = 0.20$. What is the net force exerted on the package?



Problem 3.3:
A spring is compressed 1 meter with a 3 kg block. The spring has a spring constant of 3 N/m. What is the acceleration rate of the block when it is released?

Problem 3.4:
A basket of bricks and a weight are connected by a rope that passes over a massless frictionless pulley as shown below. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. What is the tension of the rope?

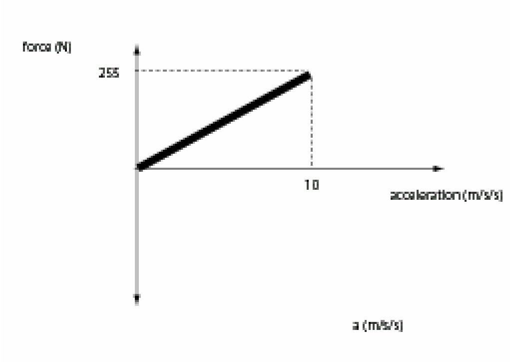


Problem 3.5:
A baseball player strikes out. In order to throw the bat at the pitcher h accelerates the bat at a rate of 10 m/s/s. The bat has a mass of 25 kg. What force was needed to accelerate the bat?

Problem 3.6:
A 100 kg person is in an elevator moving upward at 2 m/s/s. What force does the floor exert of the person?

Problem 3.7
A car has a mass of 710 kg. It starts from rest and travels 40.0m in 3.0s. The car is uniformly accelerated during the entire time. What net force is acting on the car?

Problem 3.8
An object is dragged across a floor with several different forces. The accelerations produced by these forces were recorded. A force vs. acceleration graph was produced as seen below. The last force utilized was 10 Newtons and it produced an acceleration was 5 m/s/s. What is the mass of the object?

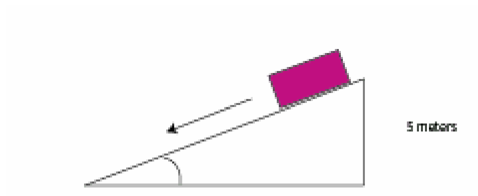


Problem 3.9
A 4600 kg helicopter accelerates upward at 2.0 m/s^2. What lift force is exerted by the air on the propellers?

IV. Energy Model:

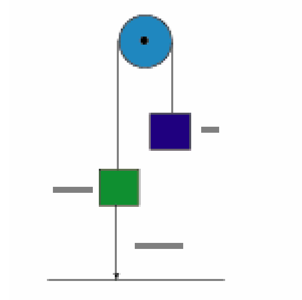
Problem 4.1
Two people named Jim and Jane jump off a freeway overpass into a lake below. Jim is twice as massive as Jane. Jim has a 14,700 Joules of energy right before he jumps. Calculate their velocities when they reach the lake's surface.

Problem 4.2:
A 2.00 kg package is released at the top of a frictionless incline. If the incline has a vertical drop of 5 meters what is its speed at the bottom of the incline?



Problem 4.3:
A spring is compressed with a 4 kg block. In its compressed state the spring has a potential energy of 4 joules. When the block is released with what velocity does it leave the spring?

Problem 4.4:
A basket of bricks and a weight are connected by a rope that passes over a massless frictionless pulley as shown below. The mass of the bricks (100 kg) is twice as great as the mass of the block. With what speed will the basket of bricks hit the ground?

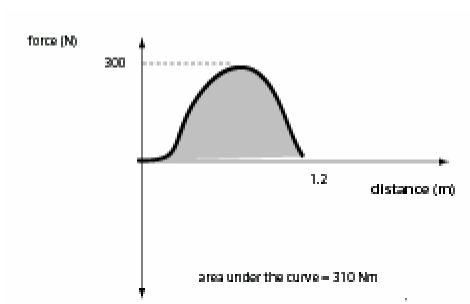


Problem 4.5:
A batter strikes out and looses the World Series. In his fury the batter throws the offending bat at a wall. The bat hits the wall with a velocity of 20 m/s and rebounds from the wall with a velocity of 18 m/s. The wall exerts, on average, a force of 5,000N opposing the motion of the bat. How far does the bat compress the wall?

Problem 4.6:
An elevator, with a mass of 500 kg, is at rest 20 meters above the ground floor on the planet Zorub that has acceleration due to gravity much lower than Earth's. The elevator has 3000 Joules of potential energy. If the elevator's cable were to snap what velocity would the elevator have right before it hits the ground floor?

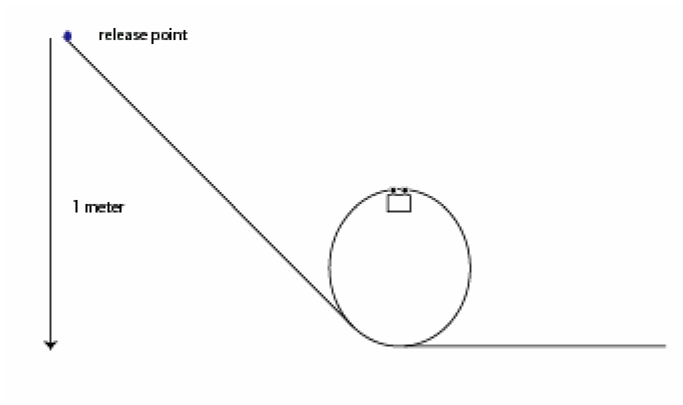
Problem 4.7:
A 2000 kg car is pulled at constant speed by a tow truck. The car moves a distance of 12.5 m across a horizontal surface. How far could the car be pulled with the expenditure of 18,000 J of work?

Problem 4.8:
A compound bow is pulled back and an arrow with a mass of 0.05 kg was inserted into the bow. A graph of the force exerted on the bow vs. the distance pulled back was plotted. If the initial velocity of the arrow is zero with what velocity does the arrow leave the bow?



Problem 4.9:
A bullet with a mass of 10 g is fired from a rifle with a barrel that is 85 cm long. Assuming that the pressures of the expanding gas to be a constant 5500 N, what speed would the bullet reach?

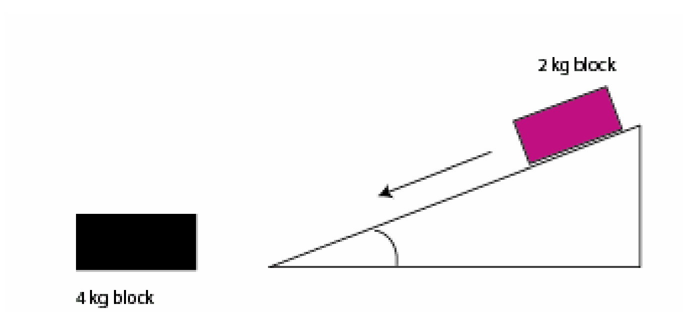
Problem 4.10:
A stunt car is released from the top of a track and runs through a loop the loop.
The initial height of release was 1 meter and the height at the top of the loop is 0.5 meters. How fast is the car moving at the top of the loop?



V. Momentum Problems:

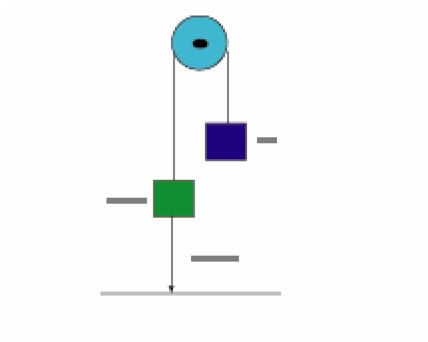
Problem 5.1:
An astronaut of mass 80 kg is in orbit freely falling around the earth. He carries an empty oxygen tank of mass 10 kg. He throws the tank away from himself with a speed of 2.0 m/s. With what speed does he start to move off into space?

Problem 5.2:
A 2.00 kg block is released at the top of an incline. The package slides down the incline and sticks to a second 4 kg block at rest at the bottom of the ramp. If the two blocks move off with a speed of 10 m/s, what was the 2.00 kg block's velocity right before it sticks to the 4.00 kg block?



Problem 5.3:
An object with mass *m* falls towards an uncompressed spring with a spring constant *k*. The instant before it encounters the spring is has a velocity *v*. After the spring is compressed by the mass it causes the mass to rebound such that the object has a velocity of *v* as it leaves the spring. What impulse did the spring experience?

Problem 5.4
A rope that passes over a mass less frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. If the system, with a total mass of 25 kg, experiences a net force of 125 Newton's over a period of 3 seconds what is the change in velocity?

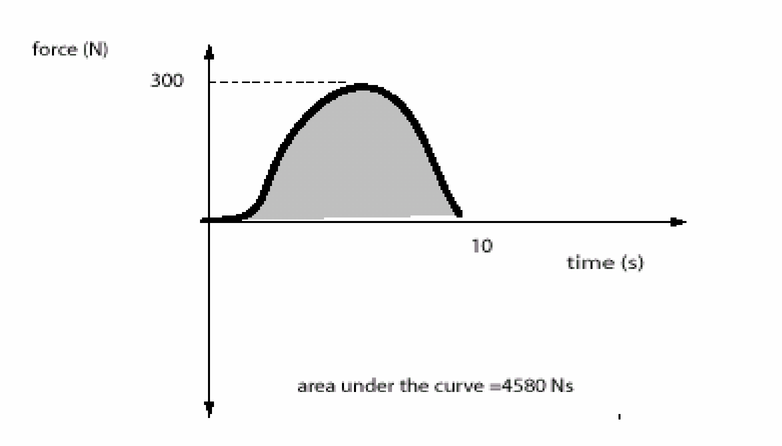


Problem 5.5
When the batter hits the ball, a net force of 1000 N, opposite to the direction of the ball's initial motion, acts on the ball for 9.0×10^{-3} s during the hit. What is the final velocity of the ball?

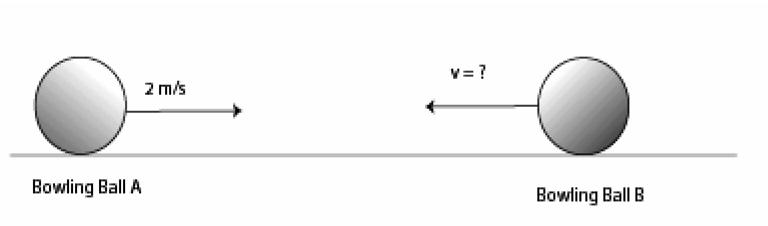
Problem 5.6
In an emergency stop an elevator is designed to collide with a large spring at the bottom of elevator shaft. If the elevator cable were to snap while the elevator is on the top floor it would reach a velocity of only 40 m/s right before encountering the stopping mechanism. The mechanism would bring the elevator to a stop after making contact with the elevator's floor for 0.50 seconds. If the elevator has a mass of 500 kg what force exerted on the elevator by the stopping mechanism?

Problem 5.7
In freight yard a train is being put together from freight cars. An empty freight car is coasting at a velocity of 10 m/s down the track. The empty freight car strikes a loaded car that is stationary, and the cars couple together. Each of the cars has a mass of 2000 kg when empty, and the loaded car contains 10,000 kg of dog food. With what velocity does the combination of the two cars start to move?

Problem 5.8
The graph below is a force vs. time of a model rocket launch. If the rocket's mass is 20 kg, what was its change in velocity?



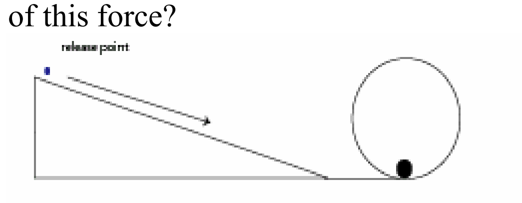
Problem 5.9
Two bowling balls with the different masses are moving towards each other. Bowling Ball A has a mass of 2.00 kg and a velocity of 2 m/s while bowling ball B has a mass of 3.5 kg. After the two bowling balls collide they have a speed of zero. How fast was bowling ball B moving before the collision?



VI. Circular Motion Model:

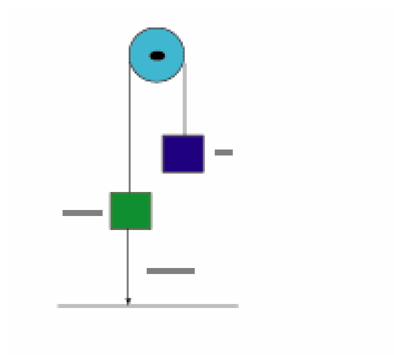
Problem 6.1:
The National Academy of Science, in order to gather information on deforestation, wishes to place a 500. Kg infrared-sensing satellite in a polar orbit (i.e., free falling) around the earth. The radius of the earth is approximately 6.38 x 103 km, and the acceleration of gravity at the orbital altitude of 160 km is very nearly the same as it is at the surface of the earth. How long would you predict it to take for the satellite to make one complete revolution around the earth?

Problem 6.2:
A ball is released from the top of an inclined plane and then rolls through a loop the loop. The diameter of the loop is 0.5 meters. If the velocity of the ball at the bottom of the ramp is 10 m/s what force must the bottom of the loop exert on the ball in order to make it move up the loop? What is the direction of this force?



Problem 6.3:
A toy plane is shot from a spring-loaded device such that it turns at a constant attitude with a radius of 100 m and a constant speed of 10 m/s. What is the acceleration of the plane?

Problem 6.4
A rope that passes over a massless frictionless pulley as shown below connects a basket of bricks and a weight. The mass of the bricks is greater then the mass of the block. The basket of bricks is released from rest at a height of 10 meters. The pulley has a radius of 0.3 meters. If it makes one rotation in 1.4 seconds, how could a fly on the edge of the pulley be feeling acceleration?

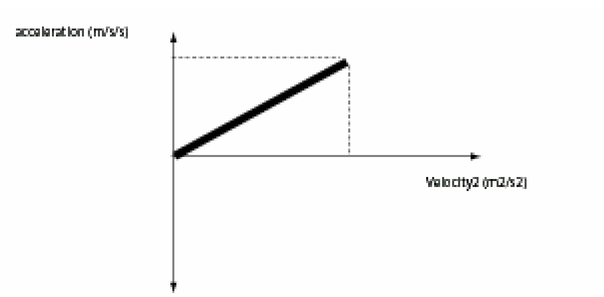


Problem 6.5
A bat with a length of 0.81 meters and a mass of 0.9 kg is swung at a ball in an arc at a constant speed of 2 m/s. The distance from the end of the handle to the center of mass is 0.4 meters. What force must be exerted on the bat in order to complete the swing?

Problem 6.6
The largest particle accelerator in the world is designed as a large circular path with a radius of 151 meters and is used to accelerate elementary particles (i.e. electrons, protons, etc.). An elevator has been specifically designed to transport scientists around the outside radius of the laboratory path with an acceleration of 10 m/s/s. If the mass of the elevator is 500 kg what is the tangential velocity felt by the elevator?

Problem 6.7
A car makes a donut in the front lawn of the next-door neighbor's house with a diameter of 10 meters. A second car makes a donut in front of another house with a diameter of 5 meters. Describe the differences in the forces and accelerations produced in the two situations.

Problem 6.8
A kids' merry – go – round is being spun at different velocities. The accelerations experienced at a constant radius were recorded for each of the velocities tested. A graph of the accelerations vs. the square of each velocity is plotted as shown below. The slope of this graph has a magnitude of 0.5. What was the constant radius?



Problem 6.9
A constant force is always applied perpendicular to the motion of a space ship, with mass m, far away from all other stellar objects. Describe the path of the ship and how you would determine the acceleration it experiences.

Problem 6.10
A toy Ferris wheel has a mass of 20 kg and a radius of 0.5 meters. If it takes the Ferris wheel 8 seconds per revolution, what is the acceleration of the outer rim of the wheel?

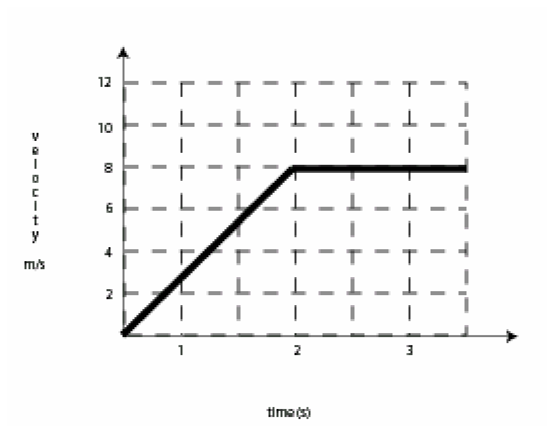
Problem 6.11
The earth's orbit around the sun is very nearly circular, with an average radius of 1.5×10^{8} km. Assume the mass of the earth is 6×10^{24} kg. What is the magnitude of the earth's average acceleration in its orbit around the sun?

APPENDIX C: Problem Solving Task (PS Task) for Study 1

Problem Solving Task
Name ______________________________

Directions: Answer the following questions to your best ability. Be sure to show all of your work.

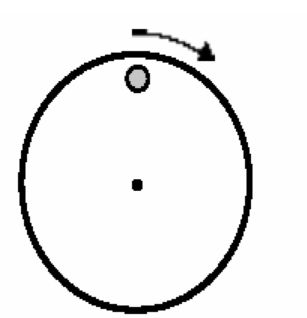
1. Using the graph below answer the following questions:



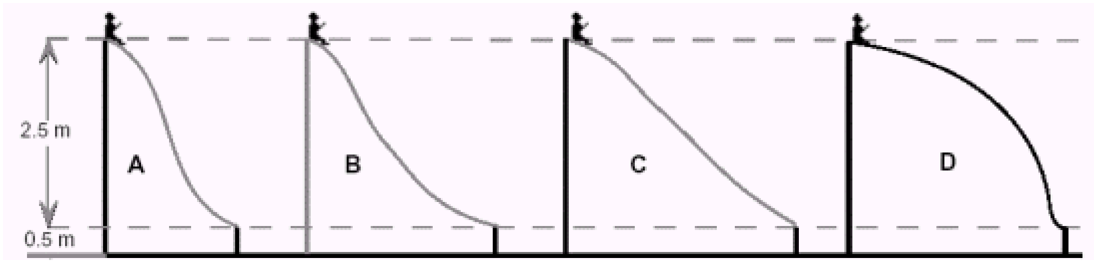
a. Calculate the acceleration for the time period displaying accelerated motion

b. Calculate the velocity during the time period displaying constant velocity.

2. A small metal cylinder with a mass of 1 kg is resting 0.5 m from the center of a circular turntable that is rotating at a constant speed (see the diagram below). It takes the cylinder 2 seconds to complete one rotation. The coefficient of static friction between the surface and the cylinder is 0.5. What is the maximum acceleration that the cylinder experiences as it moves along its circular path?



3. A young girl wishes to select the frictionless slide that will allow her to have the highest velocity at the bottom.



a. Which slide should she select and why.

b. What is her velocity at the bottom?

4. A car went 20 meters in 10 seconds.

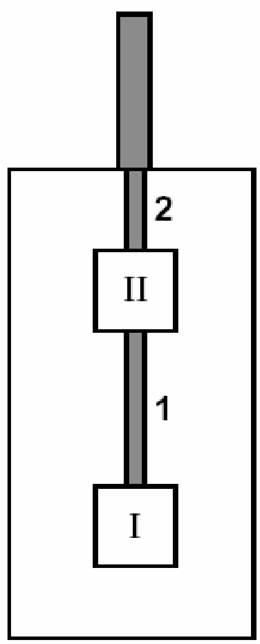
a. Assuming constant velocity calculate the average velocity.

b. Assuming constant acceleration from rest:

i. Calculate the rate at which the car changes velocity.

ii. Calculate the car's final velocity.

5. Block one and two, each with a mass of 2.0 kg, are hung from the ceiling of an elevator by ropes one and two.

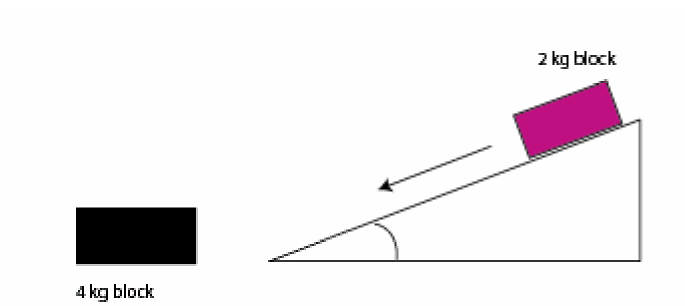


a. What is the force exerted by rope 1 on block 1 when the elevator is traveling upward at a constant acceleration of 2 m/s/s?

b. What is the force exerted by rope 2 on block II when the elevator is traveling upward at a constant acceleration of 2 m/s/s?

c. What is the force exerted by rope one on block II? Explain your reasoning?

6. A 2.00 kg block is released at the top of an incline that is 20 degrees from the horizon. The package slides down the incline and sticks to a second 4 kg block at rest at the bottom of the ramp.



a. If the two blocks move off with a speed of 10 m/s, what was the 2.00 kg block's velocity right before it sticks to the 4.00 kg block?

b. If the 2 kg block interacts with the 4.00 kg block for 0.01 seconds what was the force of the impact?

APPENDIX D: Experimenter's Script for Study 2's Verbal Protocol

Verbal Protocol Script

In this experiment I am interested in what you say to yourself as you perform the physics problems I will give you. In order to do this I will ask you to TALK ALOUD as you work on the problems. What I mean by talk aloud is that I want you to say out loud EVERYTHING that you are thinking from the time you first see the question until you give an answer. I would like you to talk constantly from the time I present each problem until you have given your final answer to the question. I do not want you to plan out what you are saying. Just act as if you are alone in the room speaking to yourself. If you are silent for any length of time I will remind you to keep talking aloud. Do you understand what I want you to do?

Good, before we turn to the real experiment, I will start with a practice problem. I want you to talk aloud while you do the problem on the whiteboard. As an example I will ask you to multiply two numbers in your head.

First read the question out loud and then continue speaking aloud while you multiply 24 times 34!

Good!

Now I would like you to solve a physics problem. I will show you a card with the physics problem. It is your task to talk aloud while you solve the problem on the whiteboard. Any questions? Please "talk aloud" while you solve the following problem!

DO PROBLEM 1; remove white board; do problem 2; remove white board; do prob. 3; remove whiteboard

Good! Now I am going to ask you some questions about what you were thinking as you solved the three problems. We are interested in what you actually can REMEMBER rather than what you think you must have thought. If possible I would like you to tell about your memories in the sequence in which they occurred while working on the question. Please tell me if you are uncertain about any of your memories. I don't want you to work on solving the problem again, just report all that you can remember about what you were thinking when solving the problem.

Show WHITEBOARD

ASK OTHER SPECIFIC QUESTIONS ABOUT PARTICULARS THAT CAME UP DURING THE TAPING SESSION.
(Example, did you think about this being a particular type of problem?) what were you thinking when you solved this problem here; what was the first thing you thought after you read the problem?; So why was that most suited to the problem?;

REPEAT WITH THE OTHER TWO WHITEBOARDS.
Then discuss personal strategies with participants.

Adapted from: *Protocol Analysis*, Erickson and Simon

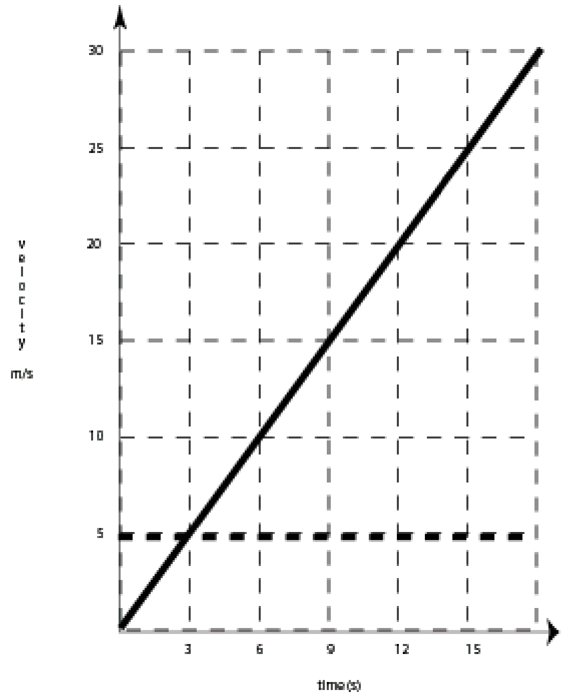
APPENDIX E: Verbal Protocol Problems for Study 2

1. Two people drop two balls off the top of the top of a tall building at the same time. Ball A has a mass of 50 kg and Ball B has a mass of 25 kg. The balls land into a bucket of water sitting on the sidewalk. If Ball A slows to a stop below the surface of the water at a rate of 5 meters per second per second what force did the water exert on the ball?

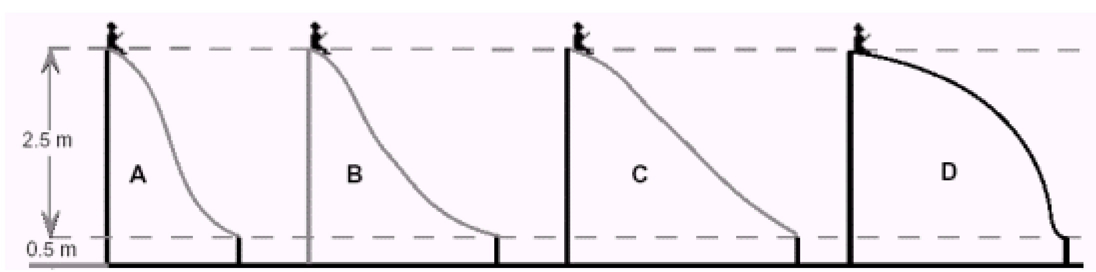
2. Two people drop two balls off the top of a tall building at the same time. Ball A has a mass of 50 kg and Ball B has a mass of 25 kg. The balls land into a bucket of water on the sidewalk. If it takes Ball A 9.5 seconds to reach the bucket how high was the building?

3. A comet has a mass of 50,000 kg and is in an elliptical orbit whose long axis is 6 X 10^{12} meters, and its period of revolution around our sun is 100 years. An asteroid has a mass of 10,000 kg and is in a circular orbit whose diameter is 2.2 X 10^{12} meters, and its period of revolution around our sun is 20 years. The comet has an x component of velocity of -5,000 m/s. It collides head on with the asteroid; whose x component of velocity is +11,000 m/s. After the collision the comet has an x component of velocity of +3,000 m/s. What is the final x component of the velocity of the asteroid?

4. A rabbit is sitting under a shade tree on a dreamy afternoon. A turtle on a super charged skateboard rolls by at a constant velocity of 5 m/s. The rabbit gives chase the instant the turtle passes him as shown in the velocity time graph below.
   a. When do the two have the same velocity?
   b. How far do they each go during the first three seconds?
   c. When and at what position does the rabbit catch up with the turtle?



5. What will be the final velocity of an object that slides down four different inclines given that the vertical height of each incline is 3.0 meters and the time it takes the objects to slide down the incline to the bottom is different for all four frictionless inclines. The first incline takes 1 second, the second incline takes three seconds, the third incline takes 4.5 seconds and the last incline takes 7 seconds.



Adapted from: Mechanics Baseline Test

APPENDIX F: Verbal Protocol Coding Scheme for Study 2

<u>Metacognition: (actively monitoring and regulating cognitive processes used to solve the problem)</u>

Planning how to approach the problem
- Setting goals
  - What type: check problem status, draw diagram, and search for equation….
- Describe solution approach in words
- "Remember" how to do this problem – (associated with problem type matching)

Monitoring comprehension
- General understanding of physics involved (self questioning, telling knowledge, Interpret graph (explain area, slope, make note))
- General understanding of the problem

Evaluating/checking (progress towards completing problem)
- Approach/path/plan
  - Does it make sense?
  - Steps in path
  - Revise plan/model selection/problem type
- Information used
- Is equation appropriate?
- Answer
  - Via alternate path
  - Recalculate
  - Does it make sense?
  - Units
  - Compare to problem - what was asked for (correct quantity?) or compare to given information
  - Sign of answer
  - Compare to problem type example
  - Revise answer

Understand/read problem:

Note given info — writing information given
Note what is asked for — writing what is asked for
Note lacking information — writing down what information is lacking
Interpreting problem givens — draw diagram while reading problem (not using it to analysis problem)>note if vector
<u>Interpreting graph and its significance – grounded in reading problem and not planning solution (i.e. id v-t graph)</u>
Note information not needed

Problem analysis: (identifying the problem and working to a solution)

<u>Problem classification</u>
- Identify model – (says something like this is a constant acceleration problem or an acceleration problem or model)
- Identify problem type - Problem examples from past mentioned – (associated with problem type matching)
- Identify Equation – (select or state equation for use in solving problem following goal for eq or looking at eq list)
  - Different types – run though a large series of equations or groups of equations without selecting one or analyzing one for possible use (i.e., type II equations)
  - Bottom up (i.e., Start with equation containing the unknown asked for in problem)
  - Forward chaining (do they start with equation without unknown variable that is asked for)

Identify algebraic representation – select a rep. from the same model initially identified – include sub headings from eq list (always this if denotes model first)

Can't Tell what led to selection – equation stated w/o reference to model, equation list etc. – include sub headings from eq list

Search textbook or equation list – [want to check if it is linked to a goal such as problem example or type]

Simplify equation – (remove variables from equation that are not needed such as when initial velocity equals zero or rearrange equation such that unknown equals all the other variables as in v = sqrt 2ad)

Expand equation – (i.e., starts with mom before = mom after and expands to $mv_1 + mv_2 = \ldots$etc.)

Combine two separate equations

Identify variables

Stating values of variables in the equation

Inputting known values for variables into equation

Lacking variable

Unnecessary variable

Inference – (guesses based upon evidence presented in the problem statement)

Note known variables not given in problem

Write known variables not given in problem

Meaning of a given variable not mentioned in problem

Diagrams

Draw diagram > What type: Picture, Free body, Motion map, Vector/direction of motion

Change or revise diagram

Graph

Draw one

Solving – get a

Algebraically

Graphically

Guess and check

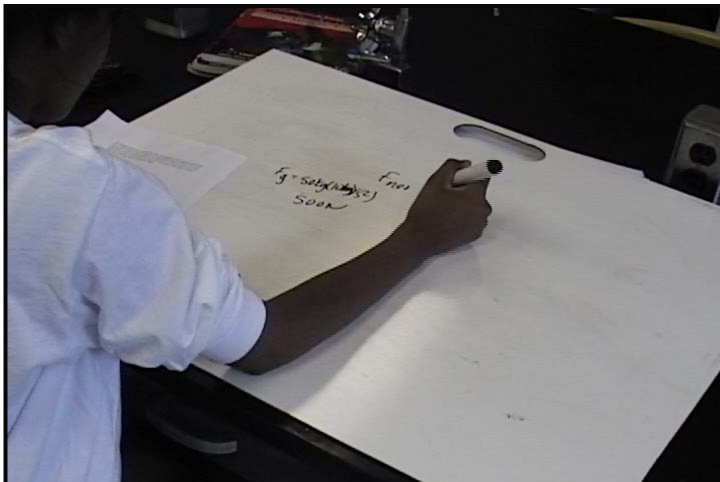
Algebraic

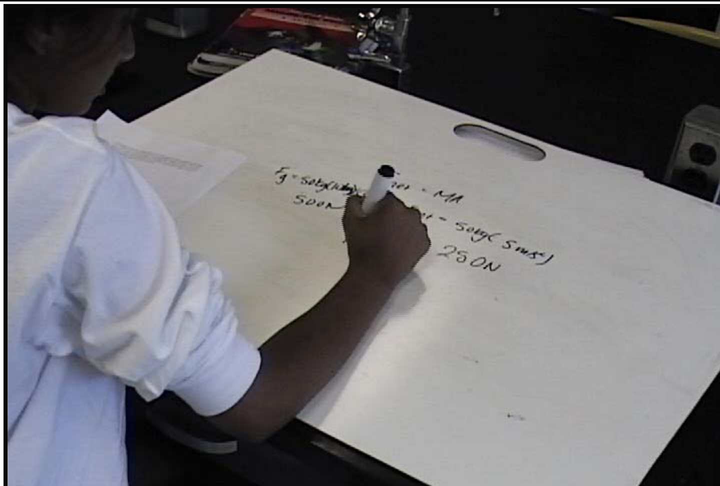
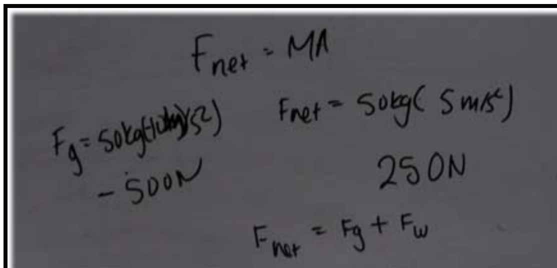
Graphical

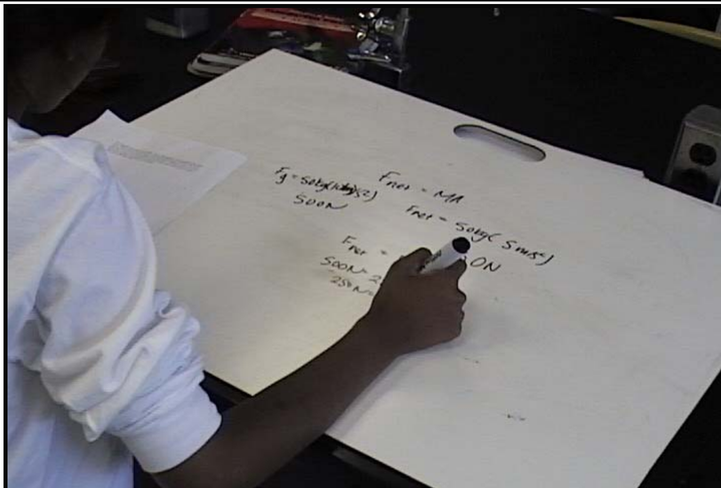
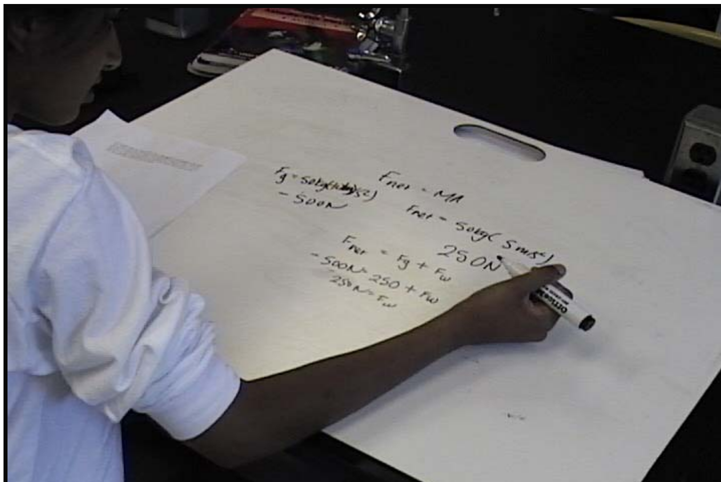
MISC

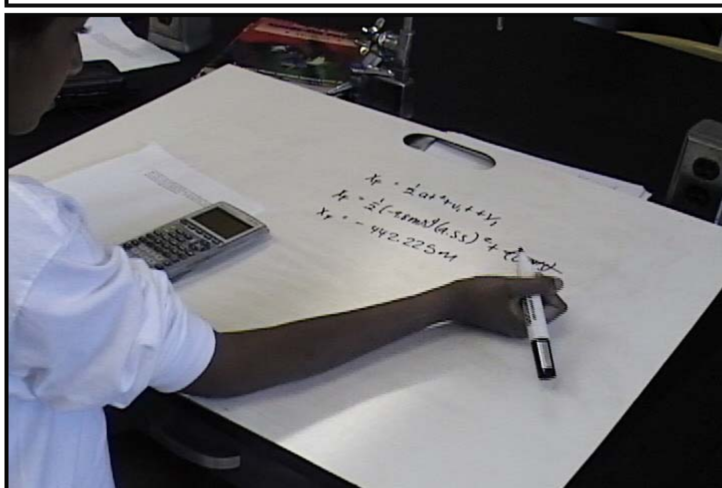
APPENDIX G: Verbal Protocol Coded Think Aloud Example from Study 2

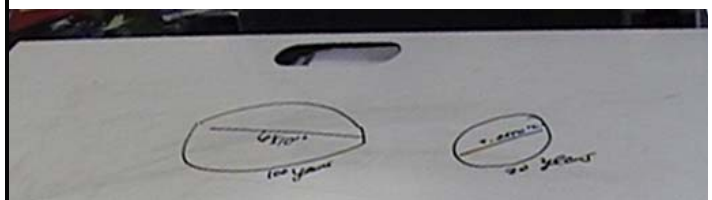
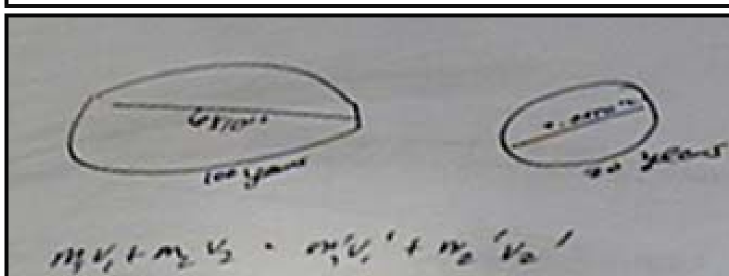
VERBAL PROTOCOL PROBLEMS –	THINK ALOUD SESSION – SUBJECT M32
PROBLEM ONE	NEWTON'S LAW 00;12;05
Two people drop two balls off the top of a tall building at the same time. Ball A has a mass of 50 kilogram and Ball B has a mass of 25 kilograms. The balls land into a bucket of water sitting on the sidewalk. If Ball A slows to a stop below the surface of the water at a rate of 5 meters per second per second what force did the water exert on the ball? [reading problem] Understand/read problem> note given info	00;35;17
Okay so I'm going to re-read the question; I have to reread it. Metacognition > planning > setting goals >reread	
Two people drop two balls off the top of a tall building at the same time. Ball A has a mass of 50 and Ball B has a mass of 25. The balls land into a bucket of water sitting on the sidewalk. Ball A slows to a stop below the surface of the water at a rate of 5 meters per second per second. [rereading problem] Understand/read problem> note given info	00;57;27
Okay so force net equals mass times acceleration [writes f net =ma] Problem analysis > problem classification > can't tell what led to the selection>forward chaining	01;00;04

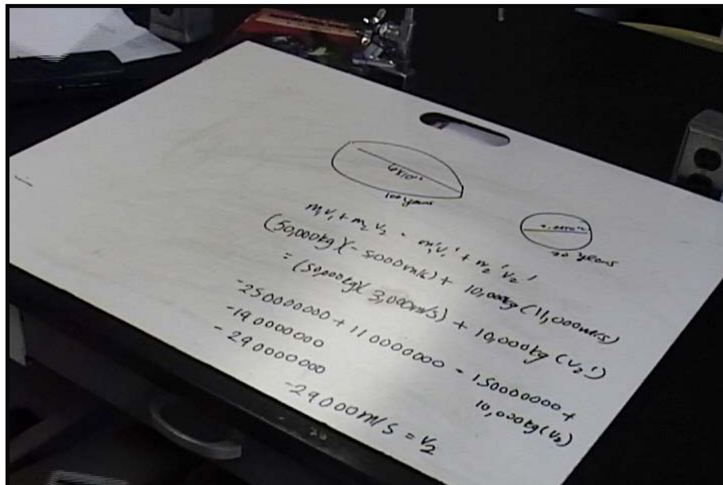
and the force that the ball is going down is 50 kilograms times 10 kilograms. [inputting values into equation] Problem analysis > id variables > inputting known values for variables into equation	
So that's… I mean meters per second squared. And so that's 500 Newtons. [solves for force of gravity] Solving > algebraically	 02;30;28
And force net equals mass 50 kilograms times 5 meters per second squared [inputting values into equation] Problem analysis > id variables > inputting known values for variables into equation	

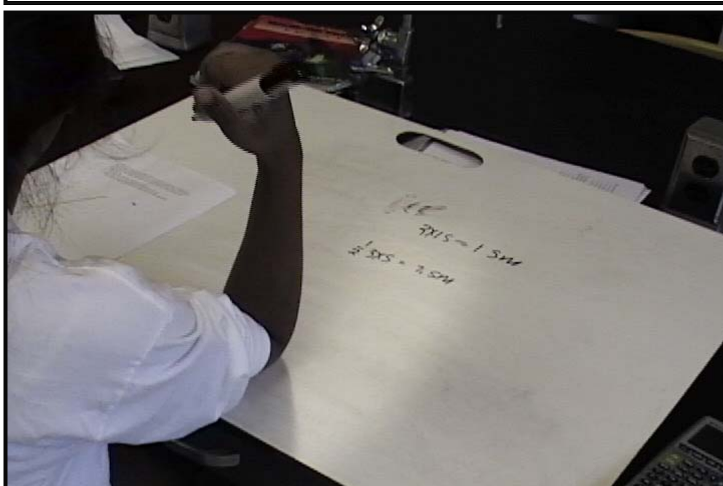
so that's 250 Newtons. [solving for F net] Solving > algebraically	 01;34;08
So the force net equals FG plus F water exerted on the ball [writes equation out] Problem analysis > problem classification > can't tell what led to the selection>forward chaining	 01;45;28
so 500 Newtons equals 250 plus FW… 7… no [inputs 500 N for the fnet and 250 for Fg and inputs values into equation] Problem analysis > id variables > inputting known values for variables into equation	

… 250 Newtons equals… force of water Solving > algebraically	 02;02;21
(*inaudible*) but I think the signs are backwards. So this is going to be negative. So 750… negative 750 Newtons. Think that's it. [solves for Fw] Metacognition > evaluating/checking > answer > sign of answer Does not catch transcription error of Fnet and Fg CORRECT	 02;10;27
PROBLEM TWO	CONSTANT ACCELERATION 02;33;29
Two people drop two balls off the top of a tall building at the same time. Ball A has a mass of 50 kilograms and Ball B has a mass of 25 kilograms. The balls land into a bucket of water on the sidewalk. If it takes Ball A 9.5 seconds to reach the bucket, how high was the building? [reading problem]	02;49;16

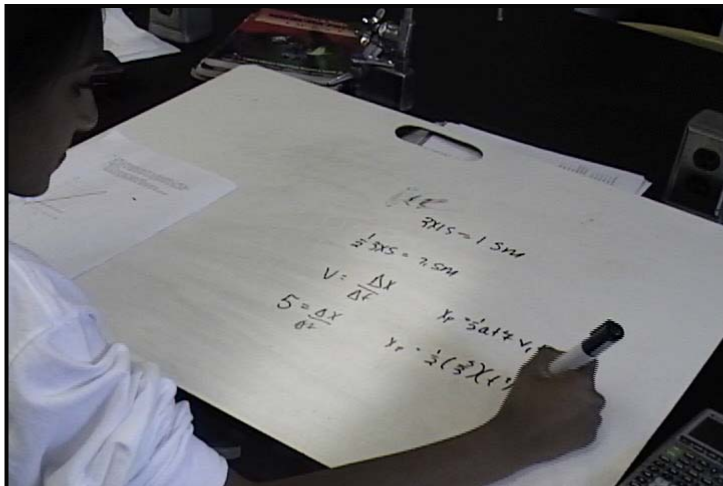
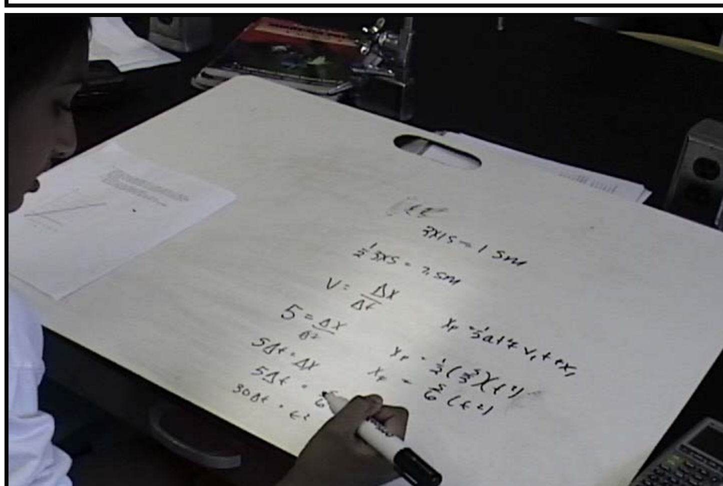
Understand/read problem> note given info	
So I'm going to use constant acceleration Problem analysis > problem classification > id model	02;50;29
[Writes out $x_f = \frac{1}{2} at^2 + v_i t + x_i$]. Problem analysis > problem classification > id algebraic representation>forward chaining	
X is 9.5 seconds. So we're trying to solve for the height ½ and the acceleration is negative 9.8 meters per second squared. And it takes 9.5 seconds… Original velocity was zero meters per second [inputting values into equation as makes inferences] Problem analysis > id variables > inputting known values for variables into equation Problem analysis > inference > write known variable not given in problem	
. So we could just cross all that off. [scratches out variables with 0] Problem analysis >simplify eq	
So X final equals…9.5 squared… Negative 442.225 meters. [uses calculator to solve] Solving > algebraically CORRECT	 03;58;18
PROBLEM 3	CONSERVATION OF MOMENTUM 04;11;19

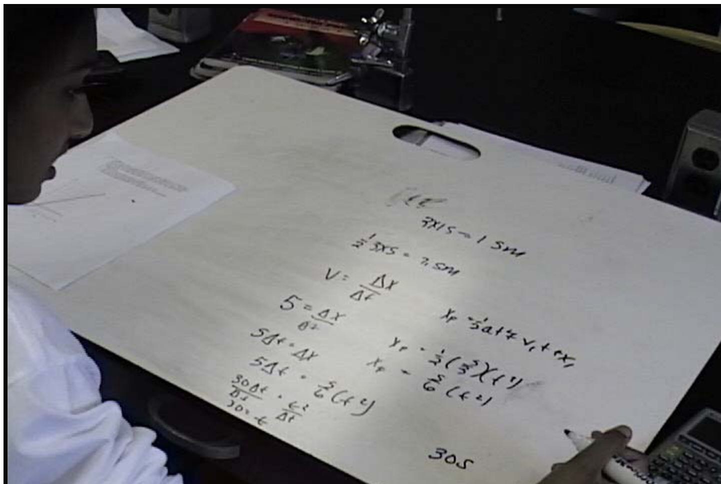
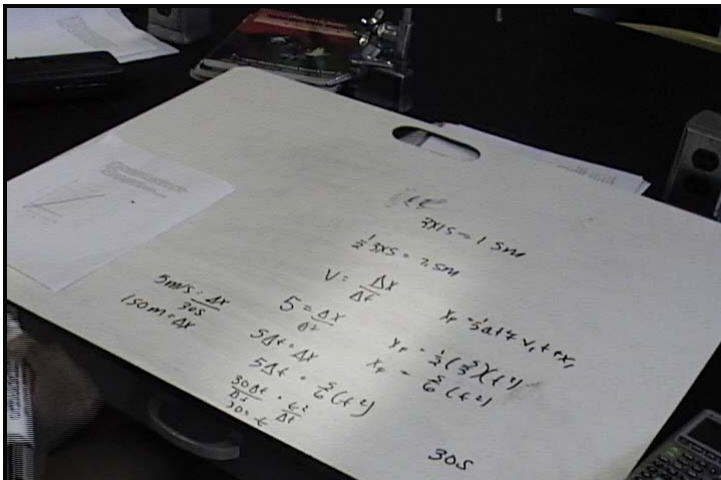
Comet has a mass of 50,000 kilograms. It's in an elliptical orbit whose long axis is 6 times 10 to the twelve meters and its period of revolution around our sun is 100 years. An asteroid has a mass of 10,000 kilograms, is in a circular orbit whose diameter is 2.2 times 10 to the twelfth meters and its period of revolution around our sun is 20 years. The comet has an x-component of velocity of negative 5,000 meters per second. It collides head-on with the asteroid whose x-component of velocity is 11,000 meters per second. After the collision the comet has an x-component of… what is the final… [reading problem and drawing a diagram of the problem labeled with known values] Understand/read problem>draw diagram Understand/read problem>writing info given	 04;59;11
oh so I'm just going to use conservation of momentum. Problem analysis > problem classification > id model	
[writes $m_1v_1 + m_2v_2 = m_1v_1' + m_2v_2'$ on board] Problem analysis > problem classification > id algebraic representation>forward chaining	 05;09;13
So the original mass was 50,000… and the velocity was… negative 5,000. And the mass of the second one was 10,000. And 11,000 meters per second equals… the masses stay the same. Comet (*inaudible*) 3,000… plus mass stays the same. [inputs known values]	

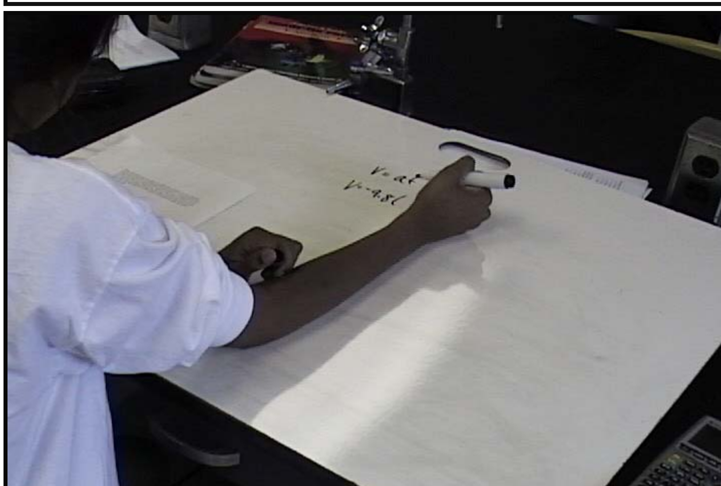
Problem analysis > id variables > inputting known values for variables into equation	
So negative 25… Negative 29000 meters per second. I think that's it. [and calculates final answer] Solving > algebraically CORRECT	 07;15;24
PROBLEM FOUR	GRAPHICAL – CONSTANT VELOCITY AND CONSTANT ACCLERATION (START TIME 07;22;24)
A rabbit is siting under a shade tree on a dreamy afternoon. A turtle on a supercharged skateboard rolls by at a constant velocity at 5 meters per second. The rabbit gives chase the instant the turtle passes him as shown in the velocity time graph below. When do the two have the same velocity? [reading problem – stops at first question] Understand/read problem>note what is asked for	07;45;01
So I'm going to find when… This is a velocity time graph so when the velocities equal Metacognition > planning > describe solution approach in words	
… at 3 seconds. [does not write it on board]	07;52;12

Solving > graphically	
And how far do they go during the first 3 seconds? Understand/read problem>note what is asked for	07;58;11
So I'm going to use displacement and during the first three seconds Metacognition > planning > setting goals >use displacement in first 3 s	08;02;27
and during the first three seconds … I don't know which one's which but the dot… the first one goes um… ½ 3… No that's not ½…it's constant [referring to graphs – id's which is which] Understand/read problem>interpreting graph	
so just 3 times 5. So 15. Solving > graphically	
And the second one goes ½ 3 times 5 so 7.5. [writing calculations on board] Solving > graphically	 08;22;19
When and at what position does the rabbit catch up with the turtle? Understand/read problem>note what is	08;26;17

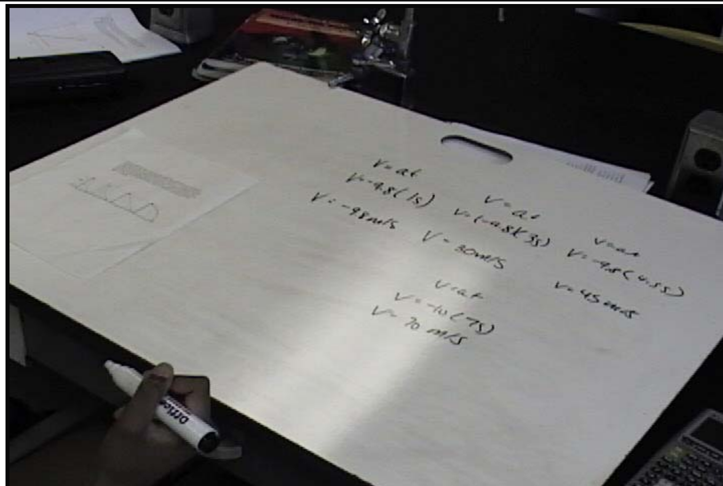
asked for	
So when is the displacements equal Metacognition > planning > describe solution path in words	08;29;23
So that would be at um… That one's 2, that one's 1… and half… 3, 4… 2… I know I didn't do that one… [becoming very quiet and referring to graph – pointing to squares] E: keep talking. Solving > graphically	09;00;09
Okay. I don't know what I'm thinking, kind of looking at it. Um… Maybe I can use… change in X… velocity equals change in X over change in time and XF equals ½ a t squared + v_i times t plus initial position. [writes equations on board] Problem analysis > problem classification > id algebraic representation>forward chaining	09;31;09
the velocity is constant Problem analysis > problem classification > id model	
and was 5. [substitutes $5 = \Delta x/\Delta t$] Problem analysis > id variables > inputting known values for variables into equation	
So the acceleration Problem analysis > problem classification > id model	

is 5/3 times t squared [substitutes in values in equation: $x_f = ½ (5/3) (t^2)$] Problem analysis > id variables > inputting known values for variables into equation	 10;00;19
Um… 5 change in T equals change in X [rearranges equation to read $5t = \Delta x$] problem analysis > simplify eq	
so 5 change in T equals 5/6 T squared. [equates the two equations] problem analysis > expand equation > combine two separate equations	
30 change in T equals T squared. problem analysis > simplify eq	 10;22;00
	10;25;01

I don't know if change in T and T are the same thing Metacognition > monitoring comprehension > general understanding of the physics	
but if they are then… 30 equals T so after 30 seconds. [solves for time] Solving > algebraically Math error > dropped a 5 from the last step and did not divide by it	 10;34;19
And to find the position that would be um 5 meters per second equals change in X over 3 seconds. So 150 meters. [solves for position using constant velocity equation] Solving > algebraically	 10;49;00
I'm not sure if that's right or not. pretend it is. Metacognition > evaluating/checking > answer CORRECT – PART ONE AND TWO INCORRECT – PART THREE – math error	10;54;27

Math error > dropped a 5 from the last step and did not divide by it	
PROBLEM 4	CONSERVATION OF ENERGY 11;01;28
What will be the final velocity of an object that slides down four different inclines given that the vertical height of each incline is 2 meters and the time it takes the objects to slide down the incline to the bottom is different for all four frictionless inclines? The first incline takes 1 second, the second incline takes 3 seconds, the third incline takes 4.5 seconds and the last incline takes 7 seconds. [reading problem] Understand/read problem> note given info	11;25;01
So I know the velocity will be… velocity is equal to AT [writes out eq on board] Problem analysis > problem classification > can't tell what led to selection>bottom up	
and the acceleration is always 9.8. [and inputs it in the equation as makes the statement] problem analysis > inference > write known variable not given in problem Problem analysis > id variables > inputting known values for variables into equation	 11;37;11
So the first one it'd take... how many seconds? 1 second. [inputs values] Problem analysis > id variables > inputting	

known values for variables into equation	
So the velocity at the bottom equals negative 9.8 meters per second. [solves for final velocity] Solving > algebraically	
And for the second one um… negative 9.8 times 3 seconds… so [inputs values] Problem analysis > id variables > inputting known values for variables into equation	
that's about… 30 meters per second. [solves for velocity at bottom of the incline] Solving > algebraically	
And V equals AT… V equals negative 9.8 times 4.5 seconds [inputs values] Problem analysis > id variables > inputting known values for variables into equation	
which is about 45 meters per second. [solves for velocity at bottom of the incline] Solving > algebraically	
And last incline takes 7 seconds so equal to AT… V equals negative 10 times 7 seconds. [inputs values] Problem analysis > id variables > inputting known values for variables into equation	
So about 70 meters per second.	

[solves for velocity at bottom of the incline] Solving > algebraically	 12;22;29
[looks over board for a bit without saying anything] I don't understand why they would each take different times though if… oh never mind, okay. So yes that's it. Metacognition > monitoring comprehension > general understanding of physics INCORRECT Physics error > using variables form different dimension in the same equation	12;39;15
RETROSPECTIVE INTERVIEW	
PROBLEM ONE –	NEWTON'S LAW
This one I was kind of confused so I had to read it again. And I think I was kind of confused because I didn't really know what Ball B had anything to do with it. But then when I tried to figure out… I'm not even sure that this one's right but… I think I screwed up the sign somewhere too	Metacognition > monitoring comprehension > general understanding of the problem

So then I just thought about the net force would have to be equal to the force that the ball was dropping with plus the force that the water was pushing back up on.	Problem analysis > problem classification >can't tell what led to selection >bottom up
E: Okay. Did you classify this as any particular type of problem? S: First I thought it was constant acceleration	Problem analysis > problem classification > id model
but then I guess once I got into it I thought it was F equals MA.	Problem analysis > problem classification >can't tell what led to selection>bottom up
E: Why did you switch from constant acceleration model to (*inaudible*)? S: I think because there's no time or anything and it just gives you the kilograms (*inaudible*). And when it asked what force did the water have so…	misc
PROBLEM TWO –	CONSTANT ACCELERATION
Well I knew that the acceleration would be 9.8 and it gave you a time was 9.5. So then you could just solve for how high it was using constant acceleration	Problem analysis > problem classification > id model
PROBLEM THREE –	CONSERVATION OF MOMENTUM
Well when I was first reading the problem I just got really confused because like the whole orbit thing. And so I decided that if it's going into orbit I don't really think… I'm not really sure if the y direction mattered.	Metacognition > monitoring comprehension > general understanding of the problem
So then I just used momentum and then I just solved that and they were real big	Problem analysis > problem classification > id model

numbers and I had to count every time. I probably should have left it in scientific notation while I was doing it.	
E: So why did you decide on momentum? S: Because it asks what… it gave you the masses and it gave you both the initial velocities and it gave you one final velocity so I thought to solve for the other one you could just do momentum.	Problem analysis > problem classification > id algebraic representation>forward chaining
PROBLEM FOUR	CONSTANT VELOCITY AND ACCLERATION WITH GRAPH
So for the first one I just saw when they matched up and that was the first problem	Solving > graphically
and B I just found the displacement and used the area under the curve	Solving > graphically
.And then for C, I didn't really know what to do so I just kind of used both of the formulas and set them equal to each other and hoped that worked.	Solving > algebraically
E: And why is this more suited for the turtle than for the Rabbit? [pointing to $v = \Delta x/\Delta t$] S: Because it's going at a constant velocity.	Problem analysis > problem classification > id model
E: Okay. And the other one? [pointing to constant acceleration equation] S: The other one goes… it's speeding	Problem analysis > problem classification > id model

up.	
E: Okay. What were you doing at first for problem C, because you sounded like you were going that's one and that's two and… S: Oh I was trying to see when the displacements would be equal so I was like counting displacements, like that one like… after three seconds it was a half [pointing to areas of the graph] and um… and 1. And then for this it was like 1½ and 2. And so I couldn't figure out when they were equal.	Solving > graphically (first)
PROBLEM FIVE:	CONSERVATION OF ENERGY
And then I guess I was thinking energy as soon as I first was reading it.	Problem analysis > problem classification > id model
But then once it said there wasn't really much about… it didn't tell you like how much… like what the mass of the person was or anything so then I decided to use V equals AT and then I just solved for them all.	Understand/read problem>not lacking info
E: Okay so first you thought it was an energy problem And then you decided to scrap that… after looking at it and decided it was a constant acceleration? S: Yes…	Problem analysis > problem classification > id model
GENERALQUESTION	GENERAL SOLUTION STRAGETIES
Well first I just kind of read the problem and I look at what it's asking usually. And then I see like what formula has all the	Problem analysis > problem classification > id equation

stuff in them… all the variables that I'm given, all the variables (*inaudible*) allowed to solve for and then I just use that one.	
E: Now how do you decide which formula to use? Do you have them like all there like in a whole big list and you just scan through them or do you… group in categories? S: Well I usually remember them so… If I see like … If I have… it like tells me like the mass and the velocity and it's asking for like the energy then I know which formula it is. So…	Problem analysis > problem classification > id model Then Problem analysis > problem classification > id alg representation>forward chaining

APPENDIX H: Physics Protocol Code Example from Study 2

Physics Protocol Code – m30

Newton's Law:

Okay so the first thing I'm going to do is I'm going to look at the actual question and since it only concerns Ball A I'm going to ignore the information they gave me about Ball B since that has nothing to do with anything.
Metacognition > planning > setting goals >ignore b [00;39;14]

So now the quantities they give me they give me acceleration, which we know as it slows to the surface so since that is… I'm going to call the downward direction negative so the upward direction will be positive. And we know that for it to slow down as it enters the water it has to… the acceleration has to be opposite to its direction of motion so since it's velocity it's downwards, its acceleration must be upwards and they give me that to be 5 meters per second per second. [draws some vectors and labels them with given values]

Metacognition > monitoring comprehension > general understanding of the problem [01;08;20]

Newton's second law valid model productive paths

↓

$ma = F_{NET}$ valid equations productive paths

↓

So now I also know that FNet equals the sum of the total forces on the ball
Metacognition > monitoring comprehension > general understanding of physics[01;23;11]

$F_{Net} = F_G + F_{water}$

↓

$F_{Net} = g(50) + F_{water}$

↓

$(50)(5) = g(50) + F_{water}$

↓

So I'll just solve now for the F_{Water},
Metacognition > planning > setting goals >what to solve for and what to expect for answer

which should be positive because we're going to let g be negative.
Metacognition > monitoring comprehension > general understanding of the problem [02;20;16]

$F_w = 250$ N

At this point I would typically check my work but that wasn't so bad so I don't think I need to but, okay. (*pause*) (*inaudible*) add in my head.
Metacognition > evaluating/checking > answer > recalculate [02;43;12]

INCORRECT – SETUP GOOD BUT ADDED INCORRECTLY
Math error > moved across equal sign forgot to change -500 to +500
Does not catch math error

Constant acceleration:

So since all we're given is a time period and we know the acceleration due to gravity – acceleration equals… Now I'm going to call the positive direction downwards just because… No I'll still call it… I'll call that negative. Downwards will still be negative. So that'll be negative 10 meters per second squared.
Metacognition > monitoring comprehension > general understanding of problem [04;22;19]

straight kinematics equation   productive path

↓

$d = \frac{1}{2} at2^{+} v_i t$ valid equation;   productive path

↓

$d = 451.25$ meters

So that's its downward displacement so if we call… So you know therefore if it fell 451.25 meters the top of the building has to be 451.25 meters off the ground. Um… yes so… building is… 451.25 meters off the ground. That doesn't really make sense but okay… 450 meters tall. The pressure's on; it's hard. Yes.
Metacognition > evaluating/checking > answer>does it make sense [05;57;12]

CORRECT

Momentum:

circular motion   nonprod path

↓   Okay so it's not circular motion because when they…
Metacognition > evaluating/checking > approach >revise model selection

not circular motion

↓   The second part of the problem is only really… Since they tell us the velocities when they collide we don't need to know the elliptical orbit or anything
Metacognition > monitoring comprehension > general understanding of the physics

conservation of momentum   prod path; valid model

And that's just because the things we're given… we're given VI and V… We're given the initial velocities of both, we're given the masses of both, we're given the final velocity of one so that leaves one component of that equation – the velocity of the other one – unsolved and therefore we can solve for it.
Metacognition > monitoring comprehension > general understanding of the problem [07;21;08]

$m_1v_1 + m_2v_2 = m_1v_3 + m_2v_4$ valid equation; productive path

And I noticed that these are both x-components, which is why it's linear change in momentum – we don't have to worry about angles or any such thing. So now we know the x-component of one. And we know that the y-components aren't going to change – they hit head on because … well it doesn't work like that.
Metacognition > monitoring comprehension > general understanding of the physics [08;15;11]

That can't… Well, I'll just go with it. 50,000 times negative 5,000 plus 10,000 times 11,000… or whatever.
[checks work]
Metacognition > evaluating/checking > answer>recalculate > section of work [09;03;18]

v_2 = -29,000 meters per second

Okay that does not seem right at all so… Uh meters per second. Let me see let's check this. 50,000 M1… M1 50,000. V1 negative 5,000… Okay so we don't know that over there. Oh yes 3,000. Is that right? Okay uh mass 2 is 10,000. Mass 2 over here is 10,000, initial velocity 11,000. minus 15… or 50 times 3… divided by 10,000. Okay well that checks off so yes this seems right then, V4 equals 29…
[checks input of known values]
Metacognition > evaluating/checking > answer >compare to given information [10;33;19]

So let's say we got a really big one moving at a decent speed, smaller one about a fifth of the size maybe at half the speed. So this one ends up increasing by that much. This one should be going… So this one switches directions. So this one… No I guess that seems… that seems reasonable.
[checks conceptual idea]
Metacognition > evaluating/checking > answer >does it make sense

So let me just check the math though. 50,000 times 3,000… So 50,000 times negative 5,000 plus 10,000 times 11,000… negative 14 negative 29,000 meters per second so it bounces off, goes the other direction. Okay.
[checks actual math calculations]
Metacognition > evaluating/checking > answer >recalculate [11;26;19]

CORRECT

Graphical problem:

a.
constant velocity valid model

↓

So they have the same velocity when the velocity of the rabbit is 5 so it's just a question of when these two lines intersect
Metacognition > planning > describe solution approach in words

3 s productive path
CORRECT

b. for turtle:
So first I'll do the turtle.
Metacognition > planning > setting goals >do turtle first

d = v*t valid equation; productive path

↓

and you can check that by saying V equals D over T and just multiply through.
Metacognition > evaluating/checking > answer>via alt path [12;37;26]

constant velocity model valid model

↓

15 m
CORRECT

and rabbit:
constant acceleration valid model

↓

which would be the area under this graph but it's much…or sorry, which would be the slope of the graph but we can also just do the integral velocity or the area under this graph to do it and that's triangle.
Metacognition > planning > describe solution approach in words [13; 07; 19]

(1/2)(3 s)(5 m/s) = 7.5 m valid equation; productive path

And yeah that does appear to be the area under the graph and we can correlate that by finding his acceleration which is going to be the slope which is rise over run so that's 5/3 meters per second per second. Now we can use D equals ½ times 5/3 times 3 seconds squared and that should equal 7.5 so 5.5 times 5 divided by 3 times 9… and that equals 7.5 so these two check off. So he doesn't get too far.
Metacognition > evaluating/checking > answer>via alt path [13;51;16]

CORRECT

c.
Ooh… Okay. So now the… Obviously they… he'll have to be moving at a faster velocity at that point.
Metacognition > monitoring comprehension > general understanding of the physics

rabbit:
$d = ½(5/3)t^2$ valid eqs; prod paths

turtle:
$= 5t$ valid eqs; prod paths

$5t = ½(5/3)t^2$ valid eqs; prod paths

$t = 6$

And now they should have the same position so this is also a convenient way of checking it.
Metacognition > monitoring comprehension > general understanding of physics [15;10;14]

$d = (5m/s)(6s)$

$= 30$ m

And now for the rabbit D equals ½ 5/3 times 6 seconds squared. That's 5/3 meters per second squared and that should equal 7.5 times 5 divided by 3 times 6 squared. Mess that up… .5 times 5 divided by 3 times 6 squared. That also equals 30 meters. So they catch up right at that… So yes, so 6 seconds and 30 meters
Metacognition > evaluating/checking > answer>via alt path [15;50;25]

6 sec and 30 meters CORRECT

Energy problem:

Now there's no way to determine accelerations or anything force… and you don't know the mass of the person so we can't use Newton's laws on these problems.
Problem analysis > problem classification > id model >different types
Metacognition > evaluating/checking > approach > does it make sense[16;56;04]

However, they do give us the vertical distance, which is kind of convenient because that allows us to determine the potential energy of the person at the top of each thing.
Problem analysis > problem classification > id model >different types
Metacognition > evaluating/checking > approach > revise [17;04;07]

potential energy — valid model; prod path

↓

$GPE = E_g = mgh$ — valid eq; prod path

$E_k = 1/2\ mv^2$ — valid eq; prod path

All right so his initial energy since he starts at rest, at rest, yes… at rest is entirely gravitational potential energy.
Metacognition > monitoring comprehension > general understanding of the physics

↓

$E_g = 20m$

All right, now then that converts to… Okay so when he reaches the 0.5 meters, the end of the slide, now since we declare that to be ground level his GPE is going to equal zero and since it's frictionless he's going to have no dissipated energy. So all this energy is KE. So KE then if he didn't lose any energy that has to be equal to his GPE initially
Metacognition > monitoring comprehension > general understanding of the physics [18;28;11]

↓

$E_k = 20\ M$

↓

$E_k = 20\ M = 1/2\ mv^2$

↓

$v = \sqrt{40}$

No square root of that. And by the way the reason I didn't make G negative is because then you get a messy negative sign here which doesn't really work too well with that V squared.
Metacognition > monitoring comprehension > general understanding of the physics

↓

$v = 6.3$ m/s — CORRECT

And that should be irrespective of the slide.
Metacognition > monitoring comprehension > general understanding of the physics [19;04;27]