MOTIVATING SCIENCE LEARNING

BY

BRIAN W. MILLER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Richard C. Anderson, Chair
Professor Fouad Abd El Khalick
Professor William F. Brewer
Professor Elizabeth A. L. Stine-Morrow
Abstract

Dual-processing theories of conceptual change hypothesize that if children are more personally involved in a lesson they will process the information more deeply leading to more and stronger conceptual change (Dole & Sinatra, 1998). This study tests this theory by increasing personal involvement through anticipation of a future discussion. Furthermore, argumentative discussions using the Collaborative Reasoning (CR) approach are thought to be more involving than regular discussions (Chinn, Anderson, & Waggoner, 2001) so exposure to CR discussions was used to increase this effect. Classrooms were randomly assigned to receive either CR discussions or regular instruction. Before reading a new text about the shape of the Earth, half of the students were informed they would later discuss the text. Students read clause-by-clause on a computer to record reading times. After reading the story, students received a comprehension test and a self-assessment survey. Before and after the intervention students were given the shape of the Earth interview (Vosniadou & Brewer, 1992) to assess their conceptual change. Students who experienced CR had longer reading times and more conceptual change but lower self-assessed depth of processing. Those students who had an announcement had increased comprehension but no other significant effects. These results partially support the dual-process theory but suggest that CR students are processing both the text and their own beliefs during reading leading to longer reading times and more conceptual change, but not improved comprehension of the text.
I dedicate this dissertation to my parents who believed in me even when I did not.
Acknowledgments

I want to acknowledge the support of Richard C. Anderson for his complete support of this project and for the Institute of Education Sciences, U.S. Department of Education (Grant # R305A080347) for financial assistance. I want to thank the members of the Collaborative Reasoning team that helped me to gather this data including Joshua Morris, Jingjing Sun, Tzu-Jung Lin, Aini Marina Ma’rof, Kim Nguyen-Jahiel, and May Jadallah. Finally, I want to acknowledge the amazing administrators, teachers, and students who were part of this project.
# Table of Contents

Chapter 1 Introduction...........................................................................................................................................1

Chapter 2 The Present Study....................................................................................................................................40

Chapter 3 Methods ..................................................................................................................................................59

Chapter 4 Results ....................................................................................................................................................94

Chapter 5 Discussion .............................................................................................................................................131

References ...............................................................................................................................................................183

Appendix A Norms of Discussion, Norms of Argument, and Instructional Moves for Collaborative Reasoning Discussions ......................................................................................................................204

Appendix B The Reading Fluency Sentence Verification Test ..................................................................................205

Appendix C The Need for Cognition Survey (Modified for Children) ................................................................209

Appendix D Shape of the Earth Interview Questions and Protocol ........................................................................211

Appendix E Shape of the Earth Interview Scoring Rubric ....................................................................................216

Appendix F A Coat for Mr. Snowman ....................................................................................................................225

Appendix G Deep Water .........................................................................................................................................228

Appendix H Transcript of a Collaborative Reasoning Discussion about the Story A Coat for Mr. Snowman ..............231

Appendix I Target Story .........................................................................................................................................242
Chapter 1

Introduction

Children enter the classrooms with many ideas about how the world works. From infancy, children begin to form expectations about the physical world. By even three months of age, many children already have an expectation that unsupported objects will fall, but supported objects will not (Baillargeon, 1994). These beliefs continue to grow and develop throughout early childhood aided by both explanations from other adults and children as well as the constant natural experimentation brought about through play and exploration (Gopnik, Meltzoff, & Kuhl, 1999). It is no surprise that by the time children are at school age, they already have many conceptions about scientific phenomena.

When children enter school they begin to learn science as a formal academic subject. Some topics in science class will concern phenomena so far outside the daily experience of children, that they will not have yet formed any ideas about them. However, for many topics students will have well-formed conceptions based on their early experiences. Students’ prior conceptions will fall a long a continuum ranging from consistent with generally excepted scientific understanding to remarkably different (Duit & Treagust, 2003).

One of the main findings of research on conceptual change is that the nature of students’ prior beliefs matters. Different educational challenges are presented and different instructional approaches are needed in each case (Carey, 2000).

When students learn a scientific concept for which they have little experience, they face the challenge of having little basis for comprehending the ideas and constructing knowledge representations from them. Forming new knowledge requires time to explore the regularities and
patterns in the world and making new connections and increasingly fine distinctions (Baillargeon, 1994).

When students try to learn a scientifically accepted theory about a phenomenon for which they already have a conception and that conception is contrary to the scientific one, then students must engage in conceptual change rather than conceptual formation. While teachers work diligently to create these changes, they are frequently unsuccessful. As Guzzetti, Snyder, Glass, and Gamas (1993) wrote in a meta-analysis of instructional interventions for conceptual change, “Whatever the rubric or the source, researchers have found that students’ misconceptions are both pervasive and remarkably resistant to change through traditional forms of instruction” (p. 118). Researchers in conceptual change have focused on what the mechanism might be that makes these conceptual changes so difficult and how the difficulty can be overcome (Duit & Treagust, 2003).

This study focuses on some aspects of how this process unfolds in classrooms. Unlike young children playing with their parents and siblings, conceptual change in school takes place in the more formal context of the classroom. This context pervades all aspects of learning including how children in school talk, read, and think. This study intends to investigate how these three processes of talking, reading, and thinking interact to support or hinder the process of conceptual change.

This view entails two contentions that will be held throughout this paper: children possess large scale coherent knowledge representations called conceptions, and conceptual change is largely intentional. The topic of conceptual change implies that there is such a thing as a conception; however the definition of a conception is not clear. As Murphy and Alexander (2008) state in their review of 20 studies on conceptual change, “One of the most central and
pervasive terms within the conceptual change literature was rarely explicitly defined and only infrequently implicitly defined. That term is ‘concept’” (p. 598). In this study, the term concept refers to a large scale knowledge representation – large scale referring to systems of interrelated information rather than isolated pieces of information (Murphy & Alexander, 2008). I will contend that these large scale representations while naïve in substance and lacking the type of formalism or institutional rigor found in the scientific community have certain similarities to scientific theories in both abstraction and coherence (Brewer & Samarapungavan, 1991).

Children’s theories, like scientific theories, provide explanations often based on unobservable entities. These explanations are generative in that they can explain new information as well as already known data (Brewer, 2008). In the instances of concepts addressed in this study, children’s conceptions of astronomy, these theories are of a particular type called naïve models. Naïve models are a subclass of theories comprised of causal mechanisms that allow the individual to run simulations of the world in the mind’s eye (Mishra & Brewer, 2003). The assumption of theories in children is supported by Brewer (2008) through the simple but powerful observation that the theories of children, such as the theory that day and night are caused by the Sun going around the Earth, was the scientifically held theory for centuries. However, this is not to say that all children have equally coherent theories or that they have any theories about issues for which they have no background information.

Intentional cognition is contrasted with algorithmic cognition which involves automated or well learned procedures. Unintentional learning can be incidental and the person can be unaware that learning has taken place. Intentional learning is goal directed and is regulated by the learner (Bereiter & Scardamalia, 1989). This study assumes that conceptual change by its nature requires a certain amount of intentional cognition. In particular, students need to work
through the difficulties and inconsistencies in their prior conceptions in order to come to more comprehensive and coherent conceptions. Therefore, students’ goals and motivations are a key factor to whether they will consciously exert the effort needed for conceptual change (Sinatra & Pintrich, 2003).

From this logic, if students participate in an instructional strategy that motivates students to increase their mental effort to understand the lesson then it should also increase conceptual change. This study seeks to examine the instructional strategy of small group discussion as a method of creating a classroom context that motivates children toward greater mental effort and then to subsequent conceptual change. It is also hoped that this research will test the theory that an increase in mental effort is the mediating mechanism by which increased motivation creates conceptual change.

I will begin by giving a brief summary of conceptual change research and the history of the approach taken in this analysis. I will then give a detailed description of the current project.

**Conceptual Change**

The study of conceptual change has roots going back to the early years of cognitive developmental psychology and in particular the work of Jean Piaget (Piaget, 1936 trans. 1977; Vosnaidou, Vamvakoussi, & Skopeliti, 2008). Piaget developed a theory for how children develop complex ideas based on a biological model of homeostasis in which the body self-regulates in response to environmental changes. Likewise, Piaget theorized that the mind seeks to adapt to the world around it to maintain balance between cognitive structures (or schemas) and the continually changing environment. Piaget proposed that there are two basic methods in which new information about the world is acted upon by the mind. The most likely method is to “assimilate” the new information into already existing schemas. This occurs when new
information fits (or nearly) fits into these existing schemas. This process is so common that Piaget terms it the “principle of universal assimilation.” However, it is not always possible for children to assimilate new information. As Piaget describes it, the world resists the child’s attempts to assimilate, because new information simply will not conform to the child’s mental organization of the world. In these cases the child must “accommodate” or change her schema to fit the new information. Piaget explains, “There are limits and obstacles to this principle of universal assimilation. These limits are imposed upon the veracity of schemata by the external world. This resistance of the external world to assimilation exacts the modification of schemata, thus adapting them to the environment. Accommodation is therefore a tendency of the organism to compensate for resistance of the object to assimilation by creating a new alternative” (Piaget, 1936 trans. 1977, p. 216). In this quotation by Piaget, we can see the origins of many important elements of conceptual change theory. Firstly, there are two different types of learning: assimilation and accommodation. These two types are roughly equivalent to the distinction between conceptual formation and conceptual change discussed above and are roughly similar to the differences between strong and weak conceptual change that will be discussed below.

Furthermore, Piaget anticipated the findings that students are resistant to conceptual change when he placed assimilation as the default mechanism. Piaget describes assimilation as the natural tendency of people and accommodation as an adjustment forced on a person by the environment. Finally, Piaget presaged the importance of anomalous data when he declared that accommodation only occurs in the face of the “resistance of the external world.”

As strong an influence as the writing of Piaget was on the study of conceptual change, Piaget had an arguably even larger indirect influence on conceptual change research through the historian and philosopher of science, Thomas Kuhn. Kuhn’s theory of how scientific theories
changes, has many similarities to the theory of Piaget on how children change their schemas, and Kuhn wrote explicitly about the influence of Piaget’s writings on the development of his influential book *The Structure of Scientific Revolutions* (Kuhn, 1962). In Kuhn’s theory, during most of the development of a mature science, scientists engage in what Kuhn refers to as “normal science” in which research falls within a paradigm. A paradigm is an overarching theory that organizes a field of research and provides “puzzles” or problems for scientists to solve in order to extend and refine the paradigm. However, sometimes data arise that do not fit into the existing paradigm. If too much of this anomalous data accumulates, it pushes the scientific community toward a revolution in which a new paradigm is developed. The idea of normal science resembles Piaget’s notion of assimilation, while the paradigm revolution is akin to Piaget’s theory of accommodation. Once again, this introduces the idea of two types of learning. The more common type in which people simply build upon the ideas they already have, and the rare type in which people must change the ideas they already have because of overwhelming anomalous data.

While Kuhn was influenced by Piaget, Kuhn’s theory differs in some fundamental ways because it is a theory of societal change while Piaget’s theory is one of individual development. In Piaget’s theory, all children naturally progress through stages that require accommodation. In Kuhn’s theory some individual scientists never accept new theories and may not even be capable of doing so. Kuhn believed that scientists using two different paradigms might have such fundamental differences that their worldviews are incommensurate. People with different paradigms ask different questions, accept different types of evidence, and even perceive observations differently (Kuhn, 1962). For this reason, Kuhn contended that individual scientists may not convert to a new paradigm while the scientific community as a whole might do so. He
describes the process of paradigm shift as a change in “the distribution of professional allegiances” (pp. 158). Furthermore, the first people to change their minds need not do so for purely rational reasons. As Kuhn describes, “The man who embraces a new paradigm at an early stage must often do so in defiance of the evidence provided by problem-solving. He must, that is, have faith that the new paradigm will succeed with the many large problems that confront it, knowing only that the older paradigm has failed with a few. A decision of that kind can only be made on faith” (pp. 158). By invoking faith, Kuhn departs from the more rational approach of Piaget. Kuhn did not imply by faith that these decisions are made arbitrarily, but that different people are persuaded by different arguments. The first people to embrace a theory might do so for “personal and inarticulate aesthetic considerations” (pp. 158). While others may be persuaded as different amounts of evidence and explanation accumulate. This idea foreshadowed the trend of putting a larger emphasis on personal variables in conceptual change that started in the early 1990’s.

In the 1970’s and 1980’s researchers discovered that children have well developed pre-instructional conceptions about science topics (Driver, Guesne, & Tiberghien, 1985; Nussbaum & Novick, 1982). In addition, researchers such as Carey (1985) and Clement (1983) saw strong similarities between the scientific revolutions described by Kuhn and the changes that children must undergo when changing from a misconception to a scientific conception. In fact, they hypothesized that children’s conceptual development recapitulated many of the same events in the history of science. The field of conceptual change soon emerged and many versions of conceptual change theory were developed to explain how children change their large scale beliefs about the world, in what ways it is similar and different from change in the history of
science, what aspects of concepts make them easier or more difficult to change, and what instructional approaches might be most effective in creating conceptual changes.

The theory described by Posner, Strike, Hewson, and Gertzog (1982) is one of the most widely cited and is sometimes called the orthodox or traditional conceptual change model (CCM). In describing this model, Posner et al. make explicit comparisons between their work and the theories of Piaget and Kuhn. Strike and Posner maintain that children enter the classroom already having many conceptions about the natural world. Some of these conceptions are not consistent with scientifically accepted theories. In a process similar to Piaget’s assimilation and Kuhn’s normal science, Posner et al. maintain that children usually take new knowledge and connect it to existing conceptions. If some of this new knowledge does not easily fit into their existing conceptions, students will merely alter or amend their conceptions rather than replace them. If these conceptions are not consistent with scientific theories, then the child has not truly learned the science curriculum. Only when there are many anomalies do students consider changing their conceptions in a manner similar to Piaget’s accommodation or Kuhn’s revolution. Even if this state of “dissatisfaction” is reached, students will not change their conceptions unless there is a new conception available to them which has certain features or “status.” This new conception must be intelligible, plausible, and fruitful. If it has this status, then change will occur. This model could be pictured as having two critical steps following a student’s exposure to a lesson containing a new concept: (1) The student will decide if the old concept is or is not satisfactory. (2) If the old concept is unsatisfactory then the student will decide if the new concept is or is not satisfactory. Only if both steps are successful will the student change their conception.
In Posner et al. (1982) original formulation of the conceptual change model, they envisioned these judgments about the dissatisfaction of old and new concepts as occurring in a broad mental context. Strike and Posner borrowed the term “conceptual ecology” from Toulmin (1972). The conceptual ecology could be considered the entire mental world of an individual – all of their beliefs, associations, memories, etc. At this point in the development of psychology it is likely impossible for researchers to understand, catalogue, or even label of what the different parts of such a mental world consist. Faced with the same daunting task, Posner et al. called the different parts of a person’s conceptual ecology “cognitive artifacts.” Unable to give a thorough definition, Strike and Posner instead offered suggestions of cognitive artifacts that would likely be important in the process of conceptual change: anomalies, analogies, metaphors, epistemological beliefs, metaphysical beliefs, knowledge from other areas of inquiry, and knowledge of competing conceptions. Much subsequent research has focused on elements of this conceptual ecology. For example, some researches such as Hewson and Hewson (1984) investigated how knowledge of competing conceptions might influence whether anomalous data was considered sufficient to cause dissatisfaction. Other researchers investigated how the conceptual ecology of metaphors and analogies can help students understand how a new conception is more plausible and fruitful than a previously held conception (Brown & Clement, 1989).

In Posner et al.’s (1982) original theory, the idea of conceptual ecology did not include affect or motivation. Strike and Posner were not unaware of these factors, but chose to concentrate on what they termed “the rational processes.” They made this choice because they wanted a theory that explained how conceptual change occurred, and they felt that motivation
and affect only facilitate learning; they are not part of the actual learning process. In the following quotation they explain their rational:

“Our central commitment in this study is that learning is a rational activity. That is, learning is fundamentally coming to comprehend and accept ideas because they are seen as intelligible and rational. Learning is thus a kind of inquiry. The student must make judgments on the basis of available evidence. It does not, of course, follow that motivational or affective variables are unimportant to the learning process. The claim that learning is a rational activity is meant to focus attention on what learning is, not what learning depends on” (p. 212).

Strike and Posner relegated motivation and affect to mediating background variables rather than variables inherent in the phenomenon itself. In this sense, Strike and Posner took a more strictly rational approach to conceptual change than Kuhn who allowed for issues of faith. Furthermore, Strike and Posner assumed that such variables as motivation could be easily controlled by basic classroom practices and that the nature of the evidence presented to students would be the primary factor in educational outcomes.

The rational tradition of the CCM inspired a large research effort that continues until the present and is the basis for many important discoveries (Özdemir & Clark, 2007; Vosniadou, 2008). Indeed, the original article of Posner et al. (1982) has been cited over 2,700 times (as measured by Google Scholar queried on 1/25/11). Some of the research based on the CCM has emphasized why a student’s prior conceptions may seem more satisfactory than scientifically accepted conceptions. One important line of research has shown prior conceptions are often seen as more satisfactory by students than scientifically accepted conceptions, because students’ prior conceptions are congruent with students’ intuitive experiences of the world. For example, the shape of the Earth appears flat rather than round (Vosniadou & Brewer, 1992) and thrown objects appear to take on the motion of the thrower (Caramazza, McCloskey, & Green, 1981; McCloskey & Kohl, 1983). Chi and colleagues (Chi, 2005; Chi & Slotta, 1993) have further
suggested that some prior conceptions are particularly difficult to change because they are of an entirely different ontological category than the scientifically accepted conceptions.

The CCM also influenced the development of a general instructional approach called the cognitive conflict strategy (Dreyfus, Jungwirth, & Eliovitch, 1990; Hewson & Hewson, 1983; Hewson & Hewson, 1984; Posner, Strike, Hewson and Gertzog, 1982; Watson & Kopnicek, 1990). This strategy has its roots in the idea of cognitive conflict or disequilibrium proposed by Piaget (Piaget, 1927 trans. 1977) to describe the state when a child’s beliefs about the world do not correspond to their experience of the world. Likewise, the cognitive conflict instructional approach attempts to confront children with the inconsistencies between their beliefs and the nature of the world. According to Límon (2001), there are three required steps in the cognitive conflict approach. First, teachers must determine what students’ prior conceptions are, then they need to confront students with the problems inherent in their prior conceptions (making them dissatisfied). This step is often done using a sort of Socratic method of questioning, but it can also be accomplished through experiments or even demonstrations that are contrary to students expectations called discrepant events. (Nussbaum & Novick, 1982). Finally, the teacher needs to present the superior aspects of the scientifically accepted view (making the students satisfied).

**Criticisms of the CCM**

As fruitful as research in the rational tradition of the CCM has been in explaining student learning, some researchers have become dissatisfied with the theory’s ability to explain why it is so difficult for students to achieve conceptual change despite increasingly targeted instructional interventions that address the conceptual gulf between prior and scientifically accepted conceptions (Duit, & Treagust, 2003). Researchers have witnessed instances in which students
seem to have every piece of necessary knowledge, and yet they do not change their conceptions (Nieswandt, 2001, Pintrich et al. 1993).

Some researchers have even become disillusioned with cognitive conflict instruction. In a targeted review of the literature on the cognitive conflict strategy, Límon (2001) summarized the disappointing results of previous studies of the conceptual conflict strategy, “Perhaps the most outstanding result of the studies using the cognitive conflict strategy is the lack of efficacy for students to achieve a strong restructuring and, consequently, a deep understanding of the new information.” Límon cites seven studies which have reported negative results for the instructional approach. Furthermore, she cites other studies in which results were mixed or incomplete. This is not to say that cognitive conflict instruction cannot be effective, but it has not been as effective overall as would be expected for an instructional method so closely matched to the theory.

Límon (2001) has suggested that the disappointing results of cognitive conflict instruction is a result of these instructional approaches being highly sensitive to contextual factors, so that differences in instructional content are overwhelmed by the variance in the context between different classrooms, children, and teachers. Límon (2001) suggests seven factors that might explain much of the difficulties with implementing cognitive conflict instruction: meaningfulness of the conflict to the students, motivation to solve the conflict, students’ prior knowledge, students’ epistemological beliefs, students’ values and attitudes, students’ learning strategies and cognitive engagement, peer interactions, and students’ reasoning abilities. Notice that many of these factors other than prior knowledge could not be directly addressed by adjusting the content of any single lesson or even unit. Instead they depend on the feelings of the student and the amount and quality of mental effort that they exert. In addition, to
these issues, Límon (2001) and Liang and Gabel (2005) have pointed out the great variance in academic outcomes from curricula might be due to the nature of student-teacher and peer interactions that make up the classroom’s social climate. These factors were certainly not unknown to those doing research on the cognitive conflict strategy, but they assumed that such factors would only make small contributions to the variance that could be neglected. However, further research has suggested that the collective impact of these factors is too strong to ignore and must be controlled for explicitly to truly test for the effectiveness of the cognitive conflict strategy itself.

In addition to the disappointing power of instructional strategies derived from the CCM, the CCM has also been insufficient to account for the complex behaviors of conceptions that have emerged from recent research. For example, researchers have suggested that people can possess multiple concepts about the same phenomena simultaneously (diSessa, 2008; Driver et al., 1994; Duit, & Treagust, 2003; Límon, 2001; Maria, 1997; Murphy, 2007, Siegler & Opfer, 2003), can switch back and forth between conceptions (diSessa, 2008; Rosenberg, Hammer, & Phelan, 2006), and can change conceptions depending on the context (diSessa, 2008; diSessa & Sherin, 1998; Duit & Treagust, 2001; Límon, 2001; Rosenberg, Hammer, & Phelan, 2006). Furthermore, they can change their concepts in a gradual manner characterized by increasing use of scientific concepts rather than a complete and permanent replacement of one concept for another (Siegler & Opfer, 2003; Smith, diSessa, & Rochelle, 1993). None of these states are addressed by the CCM. One of the rational aspects of the CCM is that children have well considered opinions that only change if the person finds the new concept has higher status (intelligible, plausible and fruitful). Using this theory there would be no reason for a child to use
two concepts simultaneously or to fluctuate between concepts if they have good reasons to believe that one theory is superior.

As with the instructional criticisms, there are explanations for the complex behavior of concepts that go beyond how the content is delivered and instead appeal to motivational, affective and attitudinal factors. For example, Murphy (2007) describes a framework in which students can express different opinions about the same scientific topic at different times, because they have not integrated their culturally-based beliefs with their school-based knowledge. For example, in Murphy’s framework a student might know that the Earth travels around the Sun, but they might be more culturally and linguistically attached to the idea that the Sun rises and sets. If that student has not thought about the issues sufficiently, he or she might not even notice a conflict. If they think about it more deeply, they might still use the more comfortable idea when they feel it is acceptable. Murphy theorizes that the degree of integration depends not on the amount of information acquired by the student, but on the depth of engagement of the student during instruction. In Murphy’s model, only high engagement will allow students to solve inconsistencies and to make firm commitments that are broadly applied. In turn, depth of engagement is determined by motivational, affective, and social factors.

The work of Murphy and Limon to create theories of conceptual change that incorporate issues of motivation, affect and social interaction is part of a larger effort that has been informally named “warm conceptual change” (Sinatra, 2005). In the next section, I will briefly introduce the history and rational of this research agenda. While this work is quite varied and has not yet settled into well-developed theoretical camps, I will also introduce what can be seen as two broad categories in the research.
Warm Conceptual Change

Through the 1980’s data began to accumulate that led to an undercurrent of criticism of the CCM introduced in the previous section. To their credit, Strike and Posner (1992) acknowledged these criticisms and published a revision of the original CCM. While the broad outlines of the theory remained, Strike and Posner expanded the theory and changed its focus. They stated that the CCM “was overly rational” (p.147). This led them to suggest that “a wider range of factors needs to be taken into account in attempting to describe a learner’s conceptual ecology. Motives and goals and the institutional and social sources of them need to be considered” (p.148). This was one of the first major statements that motivational, affective and social forces were too important to place outside of a comprehensive conceptual change theory.

Strike and Posner’s revision was amplified by Pintrich, Marx, and Boyle (1993) who described the CCM as a cold conceptual change model, because it relies primarily on describing how prior knowledge affects the processing of new information. Pintrich et al. complained that these theories do not explain how someone can posses the appropriate prior knowledge and yet not activate that knowledge at the proper times. The classic film, “A Private Universe” (Harvard-Smithsonian Center for Astrophysics, 1987), is a potent example in which a student graduating with a degree in physics from Harvard still has a naïve explanation for the seasons. Such a student undoubtedly already knows much of the disconfirming information for his beliefs (e.g. the seasons in the Northern and Southern Hemispheres are at opposite times of the year) but has not used it. Pintrich et al. feel that there cannot be an explanation for such behavior without depending strongly on motivation.

Strike and Posner (1992) and Pintrich et al. (1993) changed the perceived importance of motivation and other warm variables on conceptual change, but they did contradict the original
statement of the CCM that warm factors like motivation and affect were mediating factors and not involved in the learning process itself; they came to feel that these mediating effects were so large that they could not be ignored. Since the publication of these articles a growing number of researchers have self-identified their research as belonging to a “warming trend” in conceptual change research (Sinatra, 2005).

Warm conceptual change theory has broadened and expanded the field of conceptual change research by reconsidering many previously ignored variables. However, this proliferation of variables has made it increasingly challenging to formulate explanatory theories and organizing principles capable of incorporating all the data associated with conceptual change. Two possible organizing principles or metaphors that show some promise of helping to further the field are the idea of conceptual/learning ecologies, and conceptual change as persuasion. These two metaphors are not in opposition to each other, but they emphasize different research questions and methods. This dissertation uses primarily a conceptual change as persuasion perspective.

Conceptual Ecologies

The idea of conceptual ecologies in conceptual change research can be traced back to Posner et al.’s (1982) use of the term. This metaphor rests on the similarity between different mental entities that make up the mind and the different living and non-living parts of an ecosystem. In the same way that different plants, animals, and natural resources interact in complex ways to produce large patterns, the many aspects of the mind interact to produce large patterns such as concepts. This approach inherently views the mind as modular with different elements interacting over time to create behaviors. Furthermore, this approach allows for a
diversity of knowledge representations each having different functions and characteristics (diSessa, 2002).

One important refinement of the metaphor was made by Abd-El-Khalick and Akerson (2004). They argue that there are two different ecological views. Using the phrasing of Posner et al. (1982), warm learning factors can change what “learning is,” or it can change what “learning depends on.” When warm factors change what learning depends on, learning remains “largely a cognitive undertaking mediated and moderated by noncognitive factors” (Abd-El-Khalick & Akerson, 2004, p. 790). The warm factors in this viewpoint are static background variables. For example, in a conceptual ecology perspective motivation would be necessary to begin a process of conceptual change, but once begun is largely irrelevant to predicting if the concept itself will change which would instead depend on the logical linkages between prior and scientific conceptions. Abd-El-Khalick and Akerson reserve the original terminology of conceptual ecology for this meaning. In contrast, Abd-El-Khalick and Akerson use the term learning ecology to refer to warm variables as part of what “learning is.” In the view of learning ecologies, both cold and warm variables are interconnected with learning outcomes, and therefore they all change and interact together during the learning process. For example, a person changing their conceptions of evolution might need to simultaneously change their mental model of inheritance as well as their feelings about evolution, religion, and some of their social relationships. A learning ecology perspective would see all of these changes as part of a unified process.

Using the conceptual ecology approach, a small number of researchers have conducted correlational studies to investigate the effects of affect (Alsop & Watts, 1997; Laukenmann, Bleicher, Fu, Glaser-Zikuda, Mayring, & von Rhoneck, 2003) and motivation (Linnenbrink &
Pintrich, 2002; Palmer, 2005; Pintrich, Marx, & Boyle., 1993; Sinatra, & Mason, 2008) on achievement in conceptual change. This has produced a complex picture of interacting variables. For example, Laukenmann et al. (2003) investigated the place of affect in the conceptual change ecology of middle school students during a unit on electricity. They found that during instruction positive emotions were significantly correlated with achievement, but they were not as important when reviewing for the exam. During study a moderate level of a negative emotion, state anxiety, was a predictor of achievement. Similarly, in a review of the literature, Sinatra and Mason (2008) describe the complex interactions of interest on conceptual change as moderated by prior knowledge. While interest focuses the attention of students toward information, it also strengthens their commitments to their initial beliefs. So interest could increase or decrease conceptual change. Other motivational constructs seem to have similar complex relationships with conceptual change outcomes. These results are consistent with the metaphor of a conceptual ecology in which many mental entities have reciprocal relationships; trying to isolate the effects of any one entity is likely to lead to complex results.

In contrast to the conceptual ecology tradition, researchers in the learning ecologies approach usually take a large scale perspective. Abd-El-Khalick and Akerson (2004), Cobern (1996), Costa (1995) and Aikenhead and Jegede (1999) have studied the effect of culture, motivation, and goals on conceptual change using the learning ecology perspective. These researchers have relied on analysis of student interviews and discourse analysis to describe how the larger learning ecology of students, which includes the students’ unique cultural context, can make science very difficult to learn. They have theorized that when science is too far removed from this context, science can appear as an alien world and learning science can feel akin to crossing into a completely different country (Aikenhead and Jegede, 1999). Following the
learning ecologies approach, a person is unlikely to accept a scientific conception if it is not consistent with their worldview and identity. For example, Abd-El-Khalick and Akerson (2004) gave pre-service teachers pre and post instruction questionnaires about the nature of science. They then used in-depth interviews of three students who showed maximal improvement and three students who showed minimal improvement to investigate warm meditational factors in their conceptual growth. The students who showed maximal improvement expressed the belief that religion and science are not in opposition, that learning about the nature of science is important and useful, and that students should have deep learning goals. Furthermore, Abd-El-Khalick and Akerson described how each of these beliefs interacted with beliefs about the nature of science, thereby suggesting that these beliefs did not simply allow learning about the nature of science, but that they might be part of one integrated system of beliefs.

The learning ecology perspective has the potential of leading to a comprehensive warm conceptual change framework. However, it also presents challenges for researchers. Since the learning ecologies perspective integrates multiple aspects of each individual, it is methodologically complicated. A researcher needs to trace the motivational and affective states as well as the conceptual growth of each student and tie the two together into some pattern of mutual causation. This approach requires a great deal of information about each student. So far, this challenge has been addressed by using interviews and extended interventions, but these methodologies do not easily allow for observations of factors that students are unable to articulate, and the time intensive nature of these techniques make repeated observations over time difficult. For this reason, many researchers in this tradition have concentrated on investigating strong and persistent factors such as values and trait motivation. However, many
other warm factors such as state motivation and affect are transient and context dependent (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998).

This study is founded on a dual-process model of warm conceptual change. The model is not necessarily inconsistent with the conceptual or learning ecologies approaches. In fact, by focusing the theoretical model on the act of cognitive processing it is hoped that the dual-processing approach will bring the issues of conceptual change to a finer scale thereby allowing methodologies from desperate fields to be brought together to more easily assess the multiple factors needed to understand warm conceptual change.

**Dual-Route Processing Models**

In addition to ecology, the second metaphor that has been used to bring some organization to warm conceptual change is the metaphor of teaching as persuasion. This metaphor rests on the similarity between teachers trying to create conceptual change and the efforts of advocates trying to create attitude change. This approach is promising because concepts and attitudes have many similar features. Firstly, both concepts and attitudes are fairly complex mental entities. Dole and Sinatra (1998) argue that both attitudes and concepts are constructed by combining smaller grained mental entities: attitudes from beliefs and concepts from knowledge. Furthermore, they suggest that both beliefs and knowledge are organized into larger scale integrated structures. As discussed above, children’s concepts about science can be theory like. Similarly some social psychologists conceive of attitudes as consisting of an integrated system of beliefs (Perloff, 2003).

The attitude construct also shares many of the complex behaviors described by critics of the CCM and the warm features described by proponents of the learning ecologies perspective. Like children’s scientific concepts, the expression of attitudes is not always consistent and can be
highly context dependent (Eagly & Chaiken, 2007). For example, in a survey people might express the attitude that they love their mothers, but the way they express that attitude may vary widely according to conditions. People can even possess contradictory attitudes simultaneously, which is referred to as ambivalence in the persuasion literature (Thompson, Zanna, & Griffin, 1995).

However, there are differences between attitudes and concepts. The primary difference is that an attitude is a global evaluation that includes a strong affective component. It always makes an emotional judgment in favor or against some entity be it a person, idea, object, etc. (Perloff, 2003). A student’s scientific concept is not usually thought of in terms of an emotional evaluation. Educators often ask whether what the student thinks is true. Students are rarely asked if they like a concept. Persuasion researchers often ask whether a person likes a product, candidate, or cause. In addition, educators emphasize what students can explicitly express during classroom assessments, whereas persuasion researchers often make use of both explicit measurements and implicit measurements such as priming (Greenwald & Banaji, 1995).

While the differences between attitudes and concepts must be kept in mind, some educational researchers have felt that the similarities are great enough to make use of theories of persuasion to help explain conceptual change (Dole & Sinatra, 1998; Murphy, 2007). Of all the theories of persuasion, conceptual change researchers have been particularly attracted to the dual-process theory of persuasion called the Elaboration Likelihood Model (ELM) (Petty, & Cacioppo, 1986). The ELM is designed to describe how individual and message variables will influence the effect of a persuasive message under different levels of elaboration (Booth-Butterfield & Welbourne, 2002).
The ELM has served as an important unifying theory in persuasion research. Through the seventies, persuasion researchers were puzzled by the complex and inconsistent ways that message variables affected persuasion. For example, source authoritativeness seemed to have capricious effects on persuasion. Cook (1969) had students read a message about the dangers of brushing your teeth more than three times a day and were told that the message either came from a high school student or a dental researcher from Stanford Medical School. Students were more convinced by the dental researcher. In Rhine and Severance (1970) subjects read a short passage about plans to raise university tuition and were told that it was either a summary of a paper written by a “Yale Professor of Educational Economics” or a “Private First Class, United States Army.” After reading the message there was no main effect for the authoritativeness of the source using a Likert scale rating. Sternthal, Dholakia, and Leavitt (1978) used a similar experimental procedure but found that the moderately authoritative source (a person pursuing a career as a consumer lobbyist) was more convincing than a highly authoritative source (a Harvard-trained lawyer). Although these studies shared a bias toward private university faculty, they had completely different results and gave different explanations for their results. Cook (1969) explained his results by suggesting that authoritative sources inhibit counterarguments. Rhine and Severance (1970) explained their results by different degrees of ego-involvement. Sternhal et al. (1978) explained their results according to when the participant was informed about the source. Similar patterns could be found for other variables including the educationally relevant variable of argument quality. That people seem to be persuaded by highly logical arguments at some times and not others is the very basis of conceptual change research, but it has also been studied by persuasion researchers (O’Keefe & Jackson, 1995) as far back as Aristotle who theorized that logic was only one element of persuasion (Perloff, 2003).
Petty, Cacioppo, and Goldman (1981) used the ELM to explain all of these results in one unified theoretical framework. While they did not address exactly why each of the articles reached the particular results, Petty et al (1981) sought to replicate the same pattern of conflicting findings by manipulating a single variable of personal involvement. In the experiment students were told they would be hearing recordings of opinions about a proposed change in university policy. The participants were told that their ratings of these recordings would be part of the information given to the administration to be used in making a decision. To manipulate the degree of personal involvement, half of the subjects were told the policy would be changed in one year and half were told it would be changed in ten years. Petty et al. then sought to vary one shallow surface level aspect of the message and one substantive thoughtful level aspect of the message. For the surface level manipulation, half of the participants were told they were hearing a report written by a high school student and half were told the report was written by a professor of education from Princeton. For the substantive level manipulation, half of the students received messages with high quality arguments backed by valid statistics and data such as standardized tests results while the other half received messages backed by personal feelings and anecdotes. Participants were asked a series of Likert items to assess their opinions before and after the message. The results showed that the expertise of the speaker increased persuasion only in the low involvement condition (when the policy change would take place in ten years). The quality of the arguments was more persuasive only in the high involvement condition (when the policy change would take place next year). Petty et al. theorized that the conflicting results of previous research was due to different levels of depth of processing of the message, and that personal involvement was one way in which to change depth of processing. These results were used to support their articulation of the ELM.
Depth of processing

The heart of the ELM is stated in its second postulate: the amount and nature of the effort that people will expend on processing a message varies along an elaboration continuum (Petty, & Cacioppo, 1986). The location of a person on this continuum is referred to as their depth of processing. Depth of processing is a construct based in the information processing tradition which envisions the brain as a type of information processing machine analogous to a computer. Like a computer which has limited amounts of RAM memory and can perform only a certain number of calculations per minute, the brain also has limited working memory and can only manipulate information at a certain speed (Miller, 1956). Craik and Lockhart (1972) proposed the theory of “depth of processing” to explain certain memory phenomena including why different orienting tasks resulted in different amounts of retention. For example, asking a participant to judge the general category of each word in a list led to much better retention of those words than counting the vowels in each word on the list. Craik and Lockhart theorized that these differences in retention were due to the greater depth of processing required by each orienting task. By depth they did not mean merely quantity. They use the analogy of visual perceptual processing in which light is first processed for physical features such as brightness and edges, then for pattern recognition, and then for meaning extraction. These steps require not only additional time and amounts of resources; they also require different types of processing. In the same way, deeper processing involves both more quantity and different types of processing.

ELM theorists built on the work of early depth of processing theorists like Craik and Lockhart (Craik & Lockhart, 1972). ELM theorists suggested that where a person’s thinking falls on the “elaboration continuum” is not only a matter of how much a person elaborates, but also the reasoning strategies that they use (Petty, Rucker, Bizer, & Cacioppo, 2004). Some strategies
such as heuristics (e.g. if this celebrity likes this brand so will I), counting the mere number of arguments, and looking at surface message features (e.g. print quality of pictures) are used when elaboration is low. These strategies are called peripheral processes. In contrast, when elaboration is high people are likely to use central processes such as analyzing the logic of arguments. Petty et al. (2004) are careful to point out that having only two labels, central and peripheral, is only a matter of convenience. In fact, there is an infinitely graded continuum of different degrees of processing along which different strategies can be ordered. Furthermore, a combination of different strategies can be used at the same time.

It might be logical to assume that since accuracy is useful people would always use the maxim amount of processing capacity they have available. However, the most accurate conclusion is not always necessary, and cognitive effort is costly, so people do not generally use this strategy. The dual processing models assume that people tend to be “cognitive misers” (Taylor, 1981); therefore they use the least amount of processing that will likely reach the level of accuracy that is good enough for the situation.

The balance between a person’s reluctance to expend cognitive effort and a person’s motivation and ability determines the amount of processing a person will expend on a particular message. On the one hand, message variables make certain messages more or less difficult to process; on the other hand, a host of different personal and situational variables give people more or less motivation to overcome this tendency. This model of cognition is not unique to the ELM. For example, in reading comprehension a similar model called standards of coherence (Stine-Morrow, Miller, Gagne, & Hertzog, 2008; van den Broek, Lorch, Linderholm, & Gustafson, 2001) is used to describe how readers exert just enough effort to maintain a certain criteria of comprehension.
Depth of processing as moderator between message variables and persuasion

The ELM states that the same variables can have different effects on attitudes depending on the degree that a person elaborates a message (Petty, & Cacioppo, 1986). Petty, Cacioppo, and Goldman (1981) is one example of this approach as applied to source credibility and argument quality. In that case, source credibility had little impact under high elaboration but the quality of the message was influential. However, under low elaboration, source credibility was the more influential variable.

Regardless of how the variables interact to lead a person to retain or change their attitudes, the ELM predicts that “attitudes created or changed by the central route will be more persistent over time, will remain more resistant to persuasion, and will exert a greater impact on cognition and behavior than will attitudes changed or created through the peripheral route” (Petty et al., 2004, p. 75). Petty et al. are not suggesting that attitude change is more likely under central processing, only that if it does occur it will be stronger and more persistent.

To better describe the theory of the ELM, I present two examples that will be described using the flow chart in Figure 1 (Petty & Cacioppo, 1986). In these examples an individual is seeing a television public service announcement suggesting that people should eat more fiber. In the first example, the protagonist uses peripheral processing while in the protagonist in the second example uses central processing.

The first person came home tired and is relaxing on the couch watching TV and eating potato chips. She sees the beginning of the public service announcement. This begins the process at the first step in the flow chart (Figure 1). In the second step she assesses her own motivation. When she sees the message she feels too tired to understand a message with content about diseases, and she does not want to change her eating habits. Her motivational state and the use of
difficult words in the commercial are cues to her to process the message peripherally (the right side of the flow chart). She does not pay that much attention but she notices how easy and relaxed the actress is when making a high fiber dinner. By the end of this commercial, this viewer feels that maybe the commercial is correct and that making high fiber meals is not as difficult as she imagined and that it is probably better for her health. She ends up putting down the potato chips, but when she makes dinner the next evening she does not think about the commercial or the issue of fiber.

The second viewer is watching television at the gym. She feels alert and is worried about her health. The beginning of the message is clearly presented, and she is able to understand the arguments in the commercial. She uses central processing (the left side of the flow chart, Figure 1). She pays careful attention to the arguments of the text and compares the information to what she already thinks about the topic. She already had favorable thoughts about fiber, but she did not consider it a priority. She realizes that before she thought low fat diets were optimal, but she sees that fiber is also important. When she visits the grocery store on her way home she buys whole wheat pasta rather than the regular pasta.

These two examples emphasize that the ELM allows for attitude change through either routes. Furthermore, it explains why the same message variables have different effects in the two situations. The comfort of the spokesperson is a factor in the first persons attitude change, but the argument about fiber as a compliment to low fat is what persuades the second person. Research inspired by the ELM has investigated a host of different message and person variables. In the next two sections, two variables will be highlighted, because they will play an important role in this study.
Involvement

One of the key personal factors that determines depth of processing is involvement. Involvement describes the relevance and importance that a person feels toward a message.

In an early meta-analysis, Johnson and Eagly (1989) described three types of involvement: outcome-relevant, value-relevant, and impression-relevant. A message is outcome-relevant if the message has practical consequences to the receiver. To apply this to an educational setting, one might imagine that a medical student would have high outcome-relevant involvement in a lesson on clinical diagnosis because this knowledge will likely change her job performance as a doctor. A message is value-relevant if it “impinges on personal value systems” (p.177). Probably the most extensively studied instance of the effects of value-relevant involvement in science education is the issue of evolution and creationism. A message is impression-relevant if it affects how the viewer will be perceived by others. For example, a student working on a group science project with friends might have high impression-relevant involvement on the issue being studied by the group, because he wants to be seen as in harmony with the other group members.

Persuasion research supports a different pattern of influence for each of the three types of involvement (Levin, Nichols, & Johnson, 2000). High outcome-relevant involvement has been shown to lead to deeper processing and to more persuasion by strong arguments and less persuasion by weak arguments. High value-relevant involvement has been shown to lead to less persuasion by both strong and weak arguments although the effect for weak arguments is more pronounced. High-value-relevant involvement can lead to biased or reduced processing. Impression-relevant involvement has been shown to increase processing but to lead to slightly less persuasion for both strong and weak arguments. Levin, Nichols, and Johnson write that this
result for impression-relevant involvement suggests that impression-relevant involvement, “leads participants to take neutral, more socially defensible postures” (p. 167). However, this effect has been tested in conditions in which the participants do not know the positions of the people in their groups. If the message receiver is concerned with the impression of people who have known opinions, especially if those opinions are polarized, then the message receiver is most likely to agree with the majority (Klaczynski, 2000; Kunda, 1990; Molden & Higgins, 2005).

**Personality**

Persuasion researchers have investigated the construct called Need for Cognition (NFC) (Petty, & Cacioppo, 1986) as a personality trait that affects depth of processing. While the term “need for cognition” suggests the drive theory origins of this construct, Petty and Cacioppo are careful to emphasize that they are interested in measuring the “intrinsic motivation” people have for elaboration. They theorized NFC to be a moderately stable personality trait. Cacioppo and Petty (1982) attribute the term to Cohen, Stotland, and Wolfe (1955). Cohen et al. (1955) designed their measure to focus on participants’ desire for clarity, but Cacioppo and Petty (1982) changed the meaning to a desire and enjoyment of cognitive effort and therefore needed a new measure.

The Need for Cognition Survey (Cacioppo and Petty, 1982) was constructed by first writing a pool of opinion statements that were promising for identifying people with a high NFC. The items were rated as personally descriptive by two groups of participants: university faculty and assembly line workers. The assumption was that university faculty would naturally have higher NFC than assembly line workers. The 34 most discriminating items were retained and given to a larger sample of university students to determine that it loads onto one factor and that it has high test-retest reliability (Cacioppo & Petty, 1982). Later Cacioppo, Petty, and Kao
(1984) gave the 34 item survey and a shorter 18 item survey to 527 participants and determined that the shorter version had almost the same reliability and that versions were highly correlated ($r=0.95$, $p<0.001$).

The short version of the Need for Cognition Survey (Cacioppo et al. 1984) has been used in over two hundred studies (Petty, Brinol, Loersch, & McCaslin, 2009). The connection to the ELM was established when it was found that people with high NFC are predicted to be more motivated to elaborate on messages and therefore will tend to have higher elaboration than people with low NFC under the same circumstances (Petty, Brinol, Loersch, & McCaslin, 2009). In addition, NFC has been found to be highly correlated with the Big-Five personality (Costa, & McCrae, 1992) factors of openness to new experiences and conscientiousness (Sadowski & Cogburn, 1997). The connection between need for cognition and education has brought the NFC construct to the attention of educational researchers (Dole & Sinatra, 1998; Mason, 2001) and in particular researchers interested in discussion (O’Keefe, 1995), but no empirical data has been gathered on this topic to my knowledge and therefore no child version of the survey has been developed.

**Description of combination theories**

Woods and Murphy (2001) describe the research areas of conceptual change and persuasion as “separated at birth” arguing that since both areas study the same question of how people change their minds, they are only divided by artificial disciplinary boundaries. Not all researchers are quite as enthusiastic. Notably, Vosnaidou (2001) has written that while analogous they differ because attitudes and concepts are not identical and education and advocacy are not identical. However, there is a growing appreciation that persuasion has much to contribute to the conceptual change field.
One contribution of persuasion research might have been to reintroduce the idea of depth of processing into the area of science education. Depth of processing as an idea has had a long history in educational research including in the area of reading research. For example, depth of processing has been used by reading researchers to explain how readers create different levels of comprehension based on their depth of processing (Anderson and Pearson, 1984; Reynolds, 2000, van den Broek, Rapp, & Kendeou, 2005). However, it was through the ELM that the idea of depth of processing has also become more prominent in the area science learning (Chinn & Brewer, 1993; Dole & Sinatra 1998; Límon, 2001; Murphy, 2007).

The natural fit between conceptual change and the ELM has led many warm conceptual change theorists to allude to ideas of depth of processing. For example Pintrich et al. (1993) wrote:

Whether a new concept is intelligible or plausible is likely to be related to the depth of processing that students engage in; if they do not cognitively engage in the task, then it is unlikely that they will be able to understand the concept in an intelligent or plausible manner. In turn, depth of processing is related, at least in part, to motivational factors, such as whether learners have more of a mastery or a performance goal orientation, level of interest, and efficacy beliefs with respect to the content area and the learning strategies to be used with the content (p. 192).

Pintrich et al. did not mention the ELM by name, but they identified its primary elements of predicting depth of processing from individual and message variables.

This new view has changed the underlying picture of the critical steps in conceptual change outlined by Posner and Strike (1982). The persuasion perspective implies that there is a critical preliminary decision that the student must make: the student will decide how much effort and what type of strategy to use for processing the message. This decision will affect not only how further decisions will be made, but also how thorough and how persistent any change will
be if it occurs. This makes the conceptual change an iterative process through the following three steps:

1. The student decides how much effort and with what strategies to process the information.
2. The student will decide if the old concept is or is not unsatisfactory.
3. The student will decide if the new concept is or is not satisfactory.

The persuasion perspective falls under the warm conceptual change research agenda, because the choice of how to process information is inherently motivational, affective, and social in nature. In addition, it describes how these motivational choices can change the conceptual change process itself by changing the strategies children might use to think about the topic and changing how strong and persistent any change will be. However, this approach does not describe how concepts change together with motivational, affective, and social factors. One explanation for this is that while attitudes have been defined as a combination of emotional, cognitive and behavioral components, concepts have been defined as cognitive entities. The ELM does not mention emotional and cognitive factors in attitude change, because these factors are assumed from other elements of persuasion research. For this theory to be imported into conceptual change research it needs the complimentary work in learning ecologies to bridge this gap. Otherwise, the ELM will not entirely change the “cold” nature of current theories of conceptual change.

The Cognitive Reconstruction of Knowledge Model

The most cited theoretical attempt to combine theories of persuasion and theories of conceptual change in a rigorous model is the Cognitive Reconstruction of Knowledge Model (CRKM) (Dole & Sinatra, 1998). The CRKM has the same basic dual process structure as the ELM. Messages (or in this case lessons) are processed along a continuum of processing depth based on the personal and motivational characteristics of the students. Two basic differences
make this a model more appropriate for educational situations. Firstly, the CRKM envisions the process of change to be an iterative process. This makes sense, because unlike commercials, speeches, and other advocacy messages, lessons are long in duration and the presentation of information is often distributed over many sessions and presented in a variety of ways. Unlike the ELM that pictures the process of persuasion as a linear process always starting with a message, in conceptual change the process can start at any point and students can seek out information as well as receive it. Secondly, the CRKM uses the term engagement rather than elaboration for depth of processing to better reflect the vocabulary of educational psychology and to emphasize the use of strategies and reflective thinking. The CRKM envisions the engagement continuum as ranging from low cognitive engagement to high metacognitive engagement. These terms emphasize that one of the key qualitative differences between the processes at either end of the engagement continuum is the presence of conscious self-reflection about one’s own thinking. Dole and Sinatra (1998) hypothesize that strong conceptual change is more likely the higher the student is on the engagement continuum with strong conceptual change being impossible at the lowest end. Still even at the highest levels of engagement, change to scientifically accepted conceptions is not guaranteed. It is a necessary but not sufficient condition for strong conceptual change.

Analogous to the ELM, the level of engagement that a student chooses to use in processing a lesson is predicted by learner and message variables. The learner characteristics include their existing conceptions and their motivation. Existing conceptions was not emphasized in the ELM, but it plays a critical role in conceptual change and is a vital part of the CRKM. The CRKM characterizes prior conceptions by their strength, coherence and commitment. Motivation expands on both the CCM’s idea of “dissatisfaction.” In the CRKM dissatisfaction with a prior
conception is only one source of motivation and is combined with how personally relevant the topic is to the student, the student’s need for cognition, and the social context of the lesson.

The message variables are characterized by their comprehensibility, coherence, plausibility, and if they are rhetorically compelling. Comprehensibility and plausibility are drawn from the CCM. They refer to the fact that a message must be understandable by the person and must have convincing evidence. Coherence is drawn from another theory of conceptual change (Thagard, 1992), and it refers to the ability of the message to explain many aspects of the topic. The last characteristic which asks if a message is rhetorically compelling is specifically added to this model to allow for peripheral processing during which a person is likely to disregard the other three characteristics and focus on the rhetorical characteristics of the message.

To better explain the features of the CRKM, an example will be illustrated using the flow chart in Figure 2 (Dole & Sinatra, 1998). This example will use the same subject of eating fiber, but this time it will be a child during a classroom health lesson rather than a television commercial. The child of interest recently returned from recess. She is comfortably sitting with four classmates at a square table. When the teacher announces that they will be learning about the importance of fiber she feels no strong emotions. She has not thought about fiber a lot so her opinions do not carry a strong commitment nor is the subject highly personally relevant. However, she did assume that fiber tastes bad and is mostly for people who are constipated. When the teacher plays a short video about the importance of fiber, she does not pay close attention, but she does notice the video is well made and that the person on the video thinks that fiber is very important. By the end of the video, the student thinks perhaps fiber is more important than she previously thought, but she does not remember any of the reasons given in the video, and she has no plans of actually eating any. The class next does an activity in which each
group of students creates a high fiber meal by picking through a stack of pictures of different foods and arranging them on a chart. Then each group must calculate the amount of fiber and write an enticing description of the meal. Feeling that maybe fiber is important, and wanting to help her friends in the group, she pays more attention to the activity. She finds picking the pictures similar to what her mom does in the grocery and she thinks that she might be able to use this to help pick foods the next time her family goes shopping. She increases her involvement and volunteers to help writing the description. By the end of the activity, she has found a menu that actually sounds delicious. She decides that eating more fiber is probably pretty easy and that she could help her mom make this meal in the future. We can see from this example that the process went through several rounds that began with peripheral processing which later became central processing. Furthermore, the students existing conceptions, the social context, and the personal relevance of the message influenced the students processing of the lesson and how influential the lesson would be on the students’ beliefs.

**Empirical tests of combined theories**

Attempts to empirically test combined theories of persuasion and conception change like the CRKM have just begun. Many of the key elements of the theory have yet to be tested in educational settings. Some research has used the CRKM and other dual processes theories to explain results, but have not sought to directly test the validity of the theory (Alexander, Buehl, & Sperl, 2001). Nussbaum and Sinatra (2003) and Alexander, Fives, Buehl, and Mulhern (2002) tried to test the theory using curriculum intervention studies.

Nussbaum and Sinatra (2003) attempted to test the CRKM using argumentation as an instructional tool. They reasoned that the process of argumentation will motivate students and give them the framework for deep elaboration which is predicted to cause conceptual change in
the CRKM. Nussbaum and Sinatra had undergraduates make predictions about falling bodies. Half of the participants were asked to come up with reasons for the opposing opinion and then asked to give their predictions again. The other half were only asked to give their predictions a second time. The students who needed to first give an argument for the opposing side were more likely to eventually decide on the scientifically accepted answer than the control group. Furthermore, the experimental group gave more complete and more accurate explanations for their answers. This is impressive given the very short intervention time, and it offers direct support for the use of argumentation. However, the study did not truly test the CRKM. Even the students in the control group could have processed the material centrally, but without the experience of thinking through the opposing side they might have been unable to make the necessary cognitive leap. In other words, this intervention may gain its power by shaping the content of students’ thoughts rather than significantly changing depth of processing. This confounding might have been avoided if the researchers had attempted to measure the degree of processing during instruction. Degree of processing is however, very difficult to measure under classroom conditions over even a one hour lesson. Alternatively, students could have been given the same curriculum, but had their processing manipulated experimentally by increasing or decreasing the personal relevance of the material.

Alexander, Fives, Buehl, and Mulhern (2002) tried to create a specific curriculum that used persuasion as a model for instruction. The curriculum taught about Galileo’s theories by describing his struggle to maintain his views in the face of opposition from other scientists and religious authorities. This curriculum was compared to a curriculum which emphasized Galileo’s theories without much mention of the contentious nature of his work. Students experiencing the new curriculum had greater gains in knowledge. However, like Nussbaum and Sinatra (2003),
Alexander et al. did not measure the degree of elaboration during instruction. It is highly likely that the new curriculum was more interesting and caused students to exert more cognitive effort. However, it is also possible the new curriculum had better explanations. From this experiment, it is impossible to distinguish between a depth of processing explanation based on the motivational factors of the new curriculum and an explanation based on a clearer presentation of the information. Unless depth of processing was measured in some way, this confound is impossible to untangle.

One researcher who has attempted to indirectly measure depth of processing is Hynd (2003). Hynd inferred depth of processing from self report in a quasi-experimental design. She combined Likert scale items and interview protocols. She found that students who changed their concepts after reading a refutational text reported being “more motivated by the usefulness of the information they were going to learn, saw learning as meeting important future goals, and engaged in more cognitive effort in learning” (p.311). Hynd infers from these self report measures that students who were more motivated processed the information centrally, were better able to grasp the arguments in the text, and were therefore more persuaded by the text to change their conceptions. Hynds’ results are consistent with the dual-process theories of conceptual change, but the connection between the students’ self-reports and their actual depth of processing was not established.

Acee and Weinstein (2010) used a novel method to indirectly measurement of depth of processing of students enrolled in a statistics course. Acee and Weinstein had half of the participating students reflect on the value of statistics to their personal goals. All students were informed that there was an optional website that gave more information but was not required for the course. More students who conducted the value-reappraisal exercises went to the website.
This is an interesting indirect measure of both motivation and depth of processing, but it does not provide a measurement during the lesson.

**Commentary on dual-process theories**

Despite the small number of studies that have been conducted to test persuasion based theories of conceptual change, it remains a promising approach. Persuasion based approaches could provide a unifying framework for the findings of the conceptual ecologies approach to warm conceptual change. Just as the ELM helped persuasion theorists make sense of the complex interactions between many message and person variables, dual process models could help explain some of the complex data produced by the conceptual ecologies approach. For example, it might help answer why certain cognitive conflict instructional strategies such as discrepant events do not always work. Dual-process models would predict that a flashy demonstration meant to create a discrepant event might trigger shallow processing. In this case, the underlying arguments would not enter into children’s thinking. Dual-process theories are fruitful in that they can generate many such hypotheses for further testing. As described earlier, the dual-processing theories have less hope of informing the learning ecologies perspective, because they derive from a field in which cognitions and emotions are already combined into the construct of attitude. This work will also need to be done with the construct of conceptions.

The CRKM has been the main theoretical attempt at creating a dual-process theory for conceptual change. The CRKM applies the ELM to educational settings by making the process iterative and focusing the continuum on engagement. The CRKM also expands on the ELM by offering a specific list of learner and message variables. Many of these variables are borrowed from the CCM. However, there is as yet very little support for the importance of this particular list of variables. While the variables characterizing the learners existing conceptions, strength,
coherence, and commitment, have good face validity, they are not grounded in any particular theory. This is also true for motivation and message variables. Motivation is a subject that already has a wealth of theoretical backing which is not reflected in this list of only four components. Likewise, there is a multitude of message variables that have been investigated by persuasion research that are not reflected in the CRKM. Perhaps, it is simply too early in the process to delineate a small definitive list of variables that will be important to conceptual change. However, this does not subtract from the overall structure of the CRKM. It simply leaves it open for much further development and refinement.

Similarly it is perhaps too early to judge the empirical testing of the CRKM. Current studies are confounded by instructional content. This problem is difficult to solve because depth of processing is very difficult to measure or even manipulate over the long time frames associated with classroom instruction. All of the studies reviewed from the persuasion literature used messages of very short length that were given in laboratory conditions. Translating this research approach to children and classroom instruction will pose considerable methodological challenges that might require technological innovations such as the methods used by Acee and Weinstein (2010).
Chapter 2

The Present Study

This experiment is designed to test the persuasion-based theories of warm conceptual change. Since the inclusion of depth of processing is one of the most distinctive aspects of these models this experiment is designed to answer the questions (1) Does classroom discussion increase depth of processing? (2) Is depth of processing a good predictor of whether conceptual change occurs? The two step causal chain of social involvement increasing processing and increased processing leading to conceptual change is a clear prediction of the theory, but previous studies of the theory have had major methodological limitations. As outlined above, some previous studies have attempted to test the theory without measuring depth of processing, or they have measured it after instruction through self-report or future behavior. Furthermore, some studies confounded the motivational power and the content of instruction. The current study attempts to improve on these past attempts by including several measures of depth of processing including an online measure gathered during the instruction. In addition, it holds the content of instruction constant between treatment groups so that only the motivation varies.

The need for some manipulation to change motivation, also allows this study to provide information about the effects of small group argumentative discussion on conceptual change. Many previous studies have focused on the content and process of discussion as an instructional approach in science education. This study extends this previous work by looking at discussion as a method of changing the overall social and motivational context of the classroom.

The general hypothesis of the study are that (1) students who are reading a text for the purpose of preparing for discussion will show greater depth of processing of the text, and (2) the
this effect will be even greater for students preparing for more socially accountable discussions.

(3) Students with greater depth of processing will have more and stronger conceptual growth.

**Overview of the Design**

The study’s structure is a pre and post test design with process data. Students began with a pretest of their conceptual understanding. Half of the classrooms then conducted small group argumentative discussions while the other groups continued their regular curriculum. Then all of the students read a new science text. Before the reading, half of the students received an announcement that they would later have a discussion based on the reading. During the reading depth of processing data was gathered in three ways: reading time, a comprehension test, and a self report of processing. Finally students received a posttest of their conceptual understanding. This design created a two by two treatment design with four groups: regular curriculum and no announcement, regular curriculum and an announcement, discussions and no announcement, and discussions and an announcement.

The announcement of a future discussion was intended to be a direct way of manipulating the motivational state of the students during reading. Since all students read the same text after the announcement, content was held constant across groups which allows for an independent test of motivation. In addition the inclusion of reading time data allows for a measurement of depth of processing during the receipt of the information. Finally, the design includes a pre as well as a post test of conceptual knowledge, which controls for some of the variability in prior knowledge across individuals and classrooms.

The next section describes the experimental manipulation and its rational and the topic that was chosen as the focus of conceptual change.
Motivational Manipulation – Anticipation of Argumentation

Social accountability such as the accountability for being part of a discussion is well established in the literature as increasing processing (see Tetlock, 1985 for review of early research). One of the early investigations of the effect of anticipation of discussion on attitude change was conducted by Chaiken (1980). She had participants read a series of arguments about a topic. Some students were informed that they would later have a group discussion on that topic. Others were told they would later have a group discussion on a different topic. Chaiken measured depth of processing by having participants write all the thoughts they had about the reading. Those students who were told they would be in a discussion listed more relevant thoughts. Those same students were more likely to change their opinion and those changes were more stable over 10 days. One theoretical explanation for why this manipulation increases engagement is that the anticipation of an impending discussion increases the impression-relevant involvement of students (Johnson & Eagly, 1989). This is also known as social accountability. It is theorized that the anticipation of argument with their peers increases students’ worries that they will appear to understand the topic, be able to express a socially acceptable cogent opinion, and respond to challenges in an articulate manner. Students’ stronger concerns lead to greater motivation, subsequently deeper processing of the text, and finally more and more persistent change.

While the effects of anticipated discussion have a long history in persuasion research, they have rarely been applied to the field of science education. This is not to say that discussion and specifically argumentation has not been investigated in the context of science education. Indeed discussion and argumentation have steadily grown in importance and emphasis in the last two decades. This has been shown by its central role in research and policy documents such as
the National Science Education Standards (NRC, 1996, 2000, 2007) which suggests that there should be a change in emphasis from “science as exploration and experiment” to “science as argument and explanation” (p113). Many articles have been published on the effects of argumentation on science learning and the proper methods for instructing students in argumentation about science topics (Bell & Linn, 2000; Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre, Rodriguez, & Duschl, 2000; Kuhn, 1993; Pontecorvo, 1987; Zohar & Nemet, 2002). However, these investigations have not usually taken a depth of processing or motivational perspective. For example, in a recent introduction to the topic of argumentation in science education, Jiménez-Aleixandre and Erduran (2008) gave five rationales for using argumentation in science instruction: 1) to help students use the “cognitive and metacognitive processes characterizing expert performance” by scientists, 2) helping students to develop critical thinking, 3) helping students develop the discourse of science, 4) helping students adopt the culture of science, 5) helping students to improve their reasoning about scientific topics and to better understand the nature of science. None of these rationales include an explicit motivational element. They certainly do not discuss the change in the motivational context of the classroom when students feel that they are responsible for expressing their views to their friends as well as demonstrating their knowledge to their teachers. These rationales also omit any mention of an interaction between argumentation and other types of learning such as basic content area reading. However, if discussion is part of the overall culture of the classroom and a regular practice, students will likely interpret all activities with this context in mind, which will likely change the way students learn in small but measurable ways.

Two treatments will be used in this experiment to test the effects of discussion on students’ depth of processing. The first is similar to the manipulation used by Chaiken (1980) --
prior to reading a passage, students will either be given simple instructions or instructions including an announcement that they will later have a discussion. The difference here is that Chaiken gave participants the idea that they would either be discussing the topic they read or another topic. This manipulation was useful because it controlled for any general effects of additional instructions. However, it required an elaborate cover story, and it required the researcher to misrepresent the purpose and methods of the experiment to the participants. Creating such a cover story is impractical in the classroom setting.

The second treatment is to vary the type of discussion that students experience: either their regular classroom discussion, or discussions using the Collaborative Reasoning approach (Chinn, Anderson, & Waggoner, 2001). This treatment was added, because, unlike laboratories, classrooms have an institutionalized format for discussion that is often teacher controlled and focused on surface level comprehension (Lemke, 1990). In particular, teachers often use a pattern in which the teacher initiates a round of discussion by giving a question, students bid to answer the question usually by raising their hands, the teacher calls on a student, the student answers the question, and finally the teacher evaluates the answer (Cazden, 2001; Lemke, 1990). In this type of discussion the teacher does the majority of speaking, the teacher controls the content, and the teacher evaluates comments. Students expecting this type of discussion are likely to expect that they will speak rarely, and they will be concerned about teacher rather than student evaluation. This classroom dynamic is likely to create less social accountability and accountability of a different sort than the small peer group discussions used by Chaiken (1980).

To address this possible concern about the type of discourse being used in classrooms discussion, a second manipulation was used in which half the teachers were explicitly taught a discussion technique called Collaborative Reasoning (CR) which uses small group argumentative
discourse. CR discussions are expected to create more social accountability than regular classroom discussion because students in this type of discussion are expected to talk much more frequently, and they are often questioned and challenged by their peers. Chinn, Anderson, and Waggoner (2001) videotaped teachers leading the same reading groups in two regular discussions and two CR discussions. Chinn et al. found that during CR discussions, students had a higher percentage of the talk than during regular discussions ($M=66.5\%$, $SD=6.1$; $M=46.9\%$, $SD=8.9$ respectively), students asked more questions per minute ($M=0.41$, $SD=0.32$; $M=0.06$, $SD=0.10$), and of the questions posed to students a higher percentage were open-ended ($M=56.3\%$, $SD=9.8$; $M=30.2$, $SD=20.4$) and a higher percentage of questions challenged students’ opinions ($M=13.3$, $SD=12.6$; $M=0.2$, $SD=0.6$). Furthermore, CR discussions have an argumentative stance which encourages children to weigh different opinions using reasons and evidence. This type of discussion is particularly suited to encourage the type of reasoning needed in science understanding (Jiménez-Aleixandre & Erduran, 2008).

The two manipulations create a two by two design that can probe the effects of the anticipation of two types of classroom discussion. The hypothesis generated from the theory is that these two manipulations will have an additive effect. As predicted by Chaiken (1980), all students who have been warned of a future discussion will have increased depth of processing. However, those students warned of a future CR discussion will have even greater depth of processing because these students will feel even greater accountability and will feel accountable for a different type of understanding.

**Collaborative Reasoning**

There are many approaches to classroom discussion that feature small group discussion, frequent student talk, and critical thinking: Experience-Text-Relationship (Au, 1979), Book Club
(Raphael & McMahon, 1994), Instructional Conversations (Goldenberg, 1992), Reciprocal Teaching (Palinscar & Brown, 1984; Rosenshine & Meister, 1994), Questioning the Author (Beck et al., 1996; McKeown, Beck, & Worthy, 1993), and Philosophy for Children (Sharp, 1995) to name a few prominent examples. However, for this experiment, the Collaborative Reasoning approach to discussion was chosen, because it encourages students to evaluate evidence and to question the veracity of their beliefs and the beliefs of others. This questioning of belief is likely to help students in the process of conceptual change consistent with both warm and traditional theories. As described by Murphy, Wilkinson, Soter, Hennessey, and Alexander (2009), each of the most widely validated discussion approaches has a stance toward what the proper topic of discussion should be and how statements should be evaluated. So called ‘efferent’ approaches such as Questioning the Author emphasize learning information from the text. Expressive approaches like Literature Circles and Book Club emphasize the reader’s personal connections with the text. Critical-analytic approaches like Collaborative Reasoning and Philosophy for Children emphasize questioning the meaning and truth value of the text and generating justified individual opinions. Furthermore, discussion approaches range in the degree of freedom and interpretive control given to students. Collaborative Reasoning takes a moderate approach in which students have most of the interpretive authority but are supervised by the teacher and are assigned a topic (Waggoner, Chinn, Yi, & Anderson, 1995). Some of the specific features of this approach that are consistent with this stance are the explicit statement of a topic which has a limited number of clear competing positions, the explicit encouragement of students to take positions, present evidence and to politely challenge the position and evidence of others, and finally the absence of any requirement for consensus.

A CR discussion begins when students read a story or other text designed to raise an
engaging ethical, policy, or scientific dilemma (Anderson, Chinn, Chang, Waggoner, & Nguyen, 1998; Clark et al., 2003; Waggoner, Chinn, Yi, & Anderson, 1995). Students meet in groups of four to eight to discuss a ‘big question’ raised by the story. The big question is framed in terms of a limited number of clear competing positions. Students decide on which position they support and then present reasons and evidence for and against these positions. Students talk freely without raising their hands, and they are responsible for their own turn taking and topic management (Chinn, Anderson, & Waggoner, 2001). However, the teacher is present, and gives some structure by such means as assigning students to groups, teaching students norms of discussion, and occasionally intervening to maintain the norms and to focus students on improving the quality of their argumentation. Students are explicitly encouraged to politely challenge the position and evidence of others, and they are not required to reach a consensus at the end of the discussion. An average discussion will last 20 minutes and will end when the teacher feels that the important issues have been explored.

**Topic – Shape of the Earth**

The topic was chosen for the intervention in this study is the shape of the Earth. The first reason this topic was chosen is that it is a mental model. As Brewer (2008) describes, “Models are a subclass of theories that use mechanical/causal mechanisms.” He specifies that mental models are active. They are mental simulations or visualizations that can be manipulated in the mind’s eye to create novel inferences. They are particularly important in scientific reasoning in which the real world phenomena are unobservable or unwieldy to manipulate physically as in the case of the Earth. The shape of the Earth is not an ideal example of a model, because it does not have a causal mechanism as you would find for example in a day/night theory (Brewer, 2008). Shape is a static characteristic. However, it still qualifies to be in this category, because people
do use the shape of the Earth to make causal inferences and thereby it becomes part of mental simulations.

It should be noted that mental models are not assumed to be always stored in long-term memory. As Vosnaidou and Brewer (1992) explain, a child’s mental model of the Earth is a “dynamic structure which is created on the spot for the purpose of answering questions, solving problems, or dealing with other situations.” While mental models are constructed as needed, they are constrained by the individual’s conceptual understanding. Furthermore, it is reasonable that as individuals construct the same model repeatedly, those models then enter long term memory and are simply recreated rather than constructed anew when they are needed. While it has not been conclusively determined to what extent children at different ages hold mental model of the Earth in long term memory, it is reasonable to assume that children at the ages used in the study (9-10 years) have thought about the topic at least on several occasions, therefore the models seen at this age group are likely to be at least partially fixed in memory.

Using a topic that is a mental model has the advantage of being at an intermediate grain size. Chi (2008) describes three grain sizes of knowledge that can be the focus of conceptual change. The smallest grain size is a single idea that can usually be well summarized in a single statement. The second grain size is mental models such as the shape of the Earth. These mental representations cannot be represented by a sentence because they need to be rich enough to simulate a real world system like the Earth. Chi (2008) uses several previous studies to roughly estimate at the mental model grain size approximately 60% of students with a flawed mental model can attain a correct model from a “relatively brief instruction from text” (pg. 69). She theorizes that this process is usually the result of an accumulation of changes at the belief level. She goes on to argue that this level of change as seen in the shape of the Earth is not possible
when the change in the mental model also requires a change at the third level or category level at which conceptual change requires an ontological shift. Heat is an example of a topic that requires an ontological shift, because students must change their conception of heat from a thing to a process in order to acquire the correct mental model of heat exchange or insulation.

Taken together, the moderate size and dynamic nature of children’s mental models of the shape of the Earth make it a useful topic for this inquiry. On the one hand, this topic is more than a mere lack of knowledge. As proven repeatedly, one cannot change a child’s mental model of the shape of the earth by merely telling the child that the Earth is a sphere, or by showing them a picture, globe or other illustration. This is because it requires the child to construct a working simulation that takes numerous interconnecting issues into account. On the other hand, the shape of the Earth is not a complex theory such as evolution or thermodynamics that requires lengthy educational experiences to understand the issues involved. The shape of the Earth is a topic that is specific enough to be assessed in a single ten to fifteen minute interview and to show at least some change over a brief period of time. For example, Hayes, Goodhew, Heit, and Gillan (2003) showed that some aspects of children’s mental models of the shape of the Earth can be changed by watching four six-minute videos.

Another reason why the topic of the shape of the Earth was chosen is that it is well studied. The long history of this research was illustrated by Brewer (2008) who cites a study by G. Stanley Hall (1883) describing children’s naïve models of the Earth. Hall’s descriptions are remarkably similar to current research. However, his research was not extended until the late twentieth century. Many researchers trace the current line of research on children’s naïve models of the Earth back to Nussbaum and Novak (1976) (see Lelliott & Rollnick, 2009; Sharp & Sharp, 2007). Inspired by the work of Piaget, Nussbaum and Novak used the clinical interview
technique to explore the beliefs of children. The questions used by Nussbaum and Novak concentrated on the issue of how things fall when released from different points on the earth. They described five notions of the earth. Each notion was a closer approximation to the scientific view. While the current list of models differs from that of Nussbaum and Novak, they discovered many of the important synthetic models and were the first to describe them in detail.

The work of Nussbaum and Novak (1976) was further developed by other researchers who also looked at cross sectional samples of children (Baxter, 1989; Nussbaum 1979; Sneider & Pulos, 1983). Some researchers such as Klein (1982), looking at Mexican-American students, and Mali and Howe (1979), looking at children in Nepal, examined the cultural influences of children’s conceptions of the Earth’s shape.

Vosniadou and Brewer (1992) made a significant contribution to the literature by developing more rigorous model identification methods that led to a set of more clearly defined models. Furthermore, they added a theoretical explanation to the development of the shape of the Earth by identifying these knowledge representations as mental models and by explaining that many of these models were synthetic models meant to coordinate initial models with scientific instruction (Lelliott & Rollnick, 2009). Through their systematic interview methods over a cross section of children, Vosniadnou and Brewer identify six prominent models (Figure 3): The disc-shaped earth (flat like a disc or pancake), the rectangular Earth (also flat but with a rectangular shape), the hollow sphere (similar to a fish bowl half filled with dirt in which people live on the inside) the dual Earth (one spherical Earth in the sky and another flat Earth on which people live), the flattened Earth (on which people live on the top of a sphere which is flat on the top and bottom), and the spherical Earth (on which people either live around the outside). In addition to these categories, some children had mixed models that had aspects of one or more other models.
Vosniadou and Brewer examined children in first, third and fifth grade and found that most first grade children had a dual Earth or mixed model, third grade students had a wide range of models, and fifth grade students predominately had either a sphere or hollow sphere model (Table 18).

Vosniadou and Brewer (1992) proposed that very young children begin by using a flat rectangular model of the Earth supported by dirt and rocks, because this idea is the most congruent with common observations. Brewer (2008) would call this an alternative naïve model, because it is constructed to satisfy the child’s observations without significant influence from science. There are three major assumptions that children hold about the Earth (Brewer, 2008) that determine the features of this model. Firstly, things that look flat are flat. This is usually a reliable characteristic and the Earth from the child’s vantage point looks flat (especially in central Illinois). Secondly, in all their experience, children have almost always found that unsupported things fall. If the Earth were unsupported, they have every reason to believe that it too would fall. Thirdly, almost all the things children have seen fall have fallen down. This direction is invariant no matter where you are, so the idea that people can live all over a round object is inconsistent with all the times they have tried to place objects on round balls.

Vosniadou and Brewer (1992) theorized that before converting to the scientifically accepted round Earth model, children often create naïve synthetic models that seek to coordinate their common observations with the instruction they receive from parents and teachers that the Earth is round. Brewer (2008) categorized the synthetic models according to the three suppositions. A disc model in which the Earth is supported by dirt and rocks looks round (circular) from above, but it allows the child to hold on to all three assumptions. This is the same for a dual earth model on which the “ground” earth that we live on is supported by dirt and rocks.
A floating disc model and a hollow sphere model require children to give up the assumption that the Earth is supported, but they can retain the idea that what looks flat is flat and that things fall down. Versions of the sphere also allow children to retain some assumptions. A spherical Earth that is flattened on top allows the child to retain the assumption that things which look flat are flat. A spherical Earth on which people only live at the top allows children to retain a belief that things fall down. Eventually, these models are no longer sufficient, and then children must abandon all three assumptions and take on the spherical model in which people live all over the surface. This is called the naïve scientific model.

Brewer (2008) compared the results of ten cross sectional studies that measured children’s models of the Earth using interview techniques. He calculated the age at which each model was most frequent. He found that the peak age at which children held each model closely followed the number of assumptions that could be retained for each model. The youngest children held to models that allowed them to keep all three assumptions (disc and dual Earth models). Intermediate age children mostly used synthetic models that allowed them to retain one or two assumptions. And the oldest participants used the naïve scientific model.

Some researchers (Nobes et al., 2003) have suggested that children do not actually have synthetic models of the shape of the Earth and that the appearance of these models is an experimental artifact of using the structured interview technique. Using a forced choice format, Nobes et al. found that children held highly fragmented and inconsistent views about the shape of the Earth before adopting the scientific model. However, Brewer (2008) has presented a body of evidence drawn from research on eye-witness testimony of children that the structured interview technique is more likely to produce valid results than forced choice formats (Siegal, Waters, & Dinwiddy, 1988). The foils in a forced choice format limit the range of models
available to the child, the appearance of the models in the foils influences the child’s answers, and the limited number of foils allows children to guess. In contrast, open ended interview procedures allow for the largest possible range of answers, follow-up questions to clarify ambiguities, and the possibility of children expressing ignorance or lack of any model. Furthermore, the results of Vosniadou and Brewer’s (1992) provide indirect evidence that children possess synthetic models. Despite the strict criteria used to assign models in the study, the majority of students (82%) presented answers that consistently fit the predictions of a model. This includes the use of a variety of generative questions that students probably had never before considered. This level of consistency and coherence suggests that in the case of the shape of the Earth, children do likely possess synthetic models. Our results show less coherence than the original study, but I attribute this to differences in the interview methods used for the study which are designed to address different research questions, and these issues will be discussed in more depth below in a later section (see pg. 65).

**Depth of Processing Measures**

Mental processing is difficult to measure. Perhaps the only current technologies that approach this dream are functional magnetic resonance imaging and event-related optical signaling. Both of these techniques require sophisticated machinery, are difficult to interpret, and are impractical for classroom-based research. Instead, educational psychologists must usually depend on indirect measurements. This experiment uses three methods: comprehension, self-assessment, and reading time. Using all three methods allows for triangulation of the results. Each of these methods is directed at different aspects of processing including an online measure in the form of reading time, an objective offline measure in the form of comprehension, and a self-reported subjective offline measure.
**Comprehension.** The logic of using a comprehension test to measure depth of processing is that comprehension is often the desired product of processing - an example of judging the tree by its fruit. Evidence for this conjecture was collected in the context of memory by Craik and his colleges in the 1970’s (Craik & Tulving, 1975; Jacoby, 1978; Jacoby, Craik, & Begg, 1979). They found that when participants were forced to exert more mental effort to remember words by being made to infer the meaning, they were more likely to remember the word. In this case, participants are choosing their level of mental effort, and the result is used to measure their effort. Alone, this measure would not be sufficient because deeper processing does not always create better comprehension, but in combination it is hoped that it will provide additional insight.

**Self-assessment.** The tradition of participants reporting their own cognitive effort has a long history in the area of human factors and human performance (Gopher & Braune, 1984). In a review of cognitive load as applied to instructional design, Sweller, van Merrienboer, and Paas (1998) found that self-report measures of cognitive effort were reliable and sensitive to small changes in cognitive load. In fact, they found self-report more reliable than physiological measures such as heart rate.

The intention of this study was to measure a broader construct than the cognitive load concept investigated by Paas (1992). As theorized by Dole and Sinatra (1998), depth of processing includes mental effort in terms of amount of “metacognitive engagement.” However, there are several self-report surveys that have been developed which are designed to measure these broader constructs in a meaningful way including the Metacognitive Awareness of Reading Strategies Inventory (Marsi) (Mokhtari & Reichard, 2002) and the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & De Groot, 1990).
The Metacognitive Awareness of Reading Strategies Inventory (MARSI) (Mokhtari & Reichard, 2002) was developed by compiling a list of 60 items based on a comprehensive review of the literature on metacognition and reading comprehension. They administered the 60 item test to a large number of students and then used exploratory factor analysis to reduce the number of items to 30. They then gave the shortened version to another large sample of students along with a self-report of reading comprehension. They found that the shortened version maintained the original structure of the longer test and that students who rated themselves as good readers scored higher on the MARSI than students who claimed to be poor readers.

The Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & De Groot, 1990) was constructed in a similar manner to the MARSI (Mokhtari & Reichard, 2002). A pool of items was constructed through a review of the relevant literature on student motivation and self-regulation. The test was administered and then exploratory factor analysis was used to identify items that were strongly loaded onto a stable set of factors. These factors were then correlated with students’ grades. All the factors were significantly correlated with course grades.

By using a combination of the MARSI and MSLQ it is hoped that a broad measurement of student self-reported level of engagement and perceived effort was be gathered.

Reading time. Using reading time as a measure of depth of processing is based on the assumption that included in the speed at which people read is time for thinking and cognition. Introspection strongly suggests that we do spend time thinking as we read, but reading is a very complex phenomenon so that precisely interpreting reading times as a measure of cognition is not obvious. Reading requires the individual to understand language without the help of contextual cues such as facial expressions and tone, while at the same time moving the eyes along the page with great precision and decoding words from written symbols. Indeed,
Anderson, Hiebert, Scott, Wilkinson (1985) describe it as similar to performing a symphony orchestra.

To better explain how reading time might reflect cognition, it is helpful to review research concerning the multiple levels of meaning constructed by people attempting to understand written discourse. While there are different proposals for the exact number and nature of these levels, one of the most influential is van Dijk and Kintsch (1983). This model theorizes that readers construct a surface code, a textbase, and situation model. The surface code consists of the meaning and syntax of specific words. The textbase is the meaning derived from the relationship across groups of words. This abstraction process takes place over units of text such as sentences, clauses, or phrases (Stine-Morrow, Miller, & Hertzog, 2006) and consists of units of meaning called propositions. The textbase is more than a mere collection of word meanings, because it is an abstraction of word meanings that no longer contains information about the specific words and word order of the text (Graesser, Millis, & Zwaan, 1997). The situation model is an understanding of what the text is about as a whole. This level no longer depends on the linguistic features of the text; it is the abstracted meaning of the text (Kintsch, 1988). The abstracted meaning of the text most likely takes different forms for different types of texts; however, they all share elements of mental models in that they can function as simulations in the mind which can generate inferences (Brewer & Lichtenstein, 1982; Zwaan & Radvansky, 1998; van Dijk, 1995).

To further complicate matters, the construction of these levels of discourse representation is not a linear process in which the surface code is abstracted into propositions which are abstracted into the situation model. Instead, the levels affect each other reciprocally (Stine-Morrow, Miller, & Hertzog, 2006). For example, the meaning of ambiguous words can be
reinterpreted based on their coherence with the evolving situation model (Rayner, Pacht, & Duffy, 1994).

How can anyone perform such a symphony? As the old joke goes, “How do you get to Carnegie Hall? Practice!” (Cerf, 1956, pg. 335). People get better and better at using their cognitive abilities to perform the simultaneous procedures needed for reading. However, even very skilled readers do not always achieve a perfect level of comprehension of text, and the level of comprehension that they reach changes according to their motivation (Stine-Marrow, Miller, Gagne, & Hertzog, 2008; van den Broek, Lorch, Linderholm, & Gustafson, 2001). In fact, readers have been shown to typically expend just enough mental resources to achieve a “good enough” understanding of the text (Christanson, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002).

There are many ways to examine the process of reading, but one of the most powerful is to analyze reading times. The time it takes to read a text can be thought of as the sum of all of the processes of comprehension discussed above. This time is influenced by the features of the text, the characteristics of the reader (e.g. motivation, fluency), and the interaction between them. Once reading times are measured, the variance can be decomposed into the many factors that contribute (Graesser, Millis, & Zwaan, 1997; Stine-Morrow, Miller, Gagne, & Hertzog, 2008). Some of the variance can be attributed to text characteristics most associated with surface level comprehension including word length (Rayner & McConkie, 1976) and the frequency of the word in the language (Inhoff & Rayner, 1986). Texts can also differ in ways that make constructing a textbase representation more or less difficult. For instance, the time needed to read a sentence increases with the number of propositions even after controlling for length (Kintsch & Keenan, 1973). Readers also spend more time at the end of meaningful units of text such as
sentences in a process termed “wrap-up.” During wrap-up, readers are likely spending additional time to integrate the information they have read into higher discourse representations (Just & Carpenter, 1980). Reading times also show the effects of constructing the situation model. For example, readers increase their speed as they continue through a text, because they are aided in their comprehension by situation model constructed from reading earlier in the text (Gernsbacher, 1990). In contrast, readers slow down when a new topic is introduced especially if this topic reflects a major shift in the argument of an expository text (Hyona, Lorch, & Kaakinen, 2002; Lorch, Lorch, Gretter, & Horn, 1987).

Taking all of these levels into account, a fairly accurate model can be constructed that significantly predicts the reading times of students with given characteristics on a particular passage. The logic of using reading times to measure the effect of discussion on depth of processing is that if discussion influences depth of processing it should have an effect on reading times in addition to the variables already in the model representing features of the language in the text and characteristics of the reader.

**Hypotheses**

The dual-route processing theories applied to conceptual change including the CRKM (Dole & Sinatra, 1998) would predict that (1) students who read for to prepare for a discussion would have greater reading times, higher self-report scores, and better comprehension of the text. (2) Students who had experience with CR discussions would have even larger results on these three measures. (3) Finally, students who had greater depth of processing would describe more sophisticated models of the shape of the Earth in their posttest interviews than during their pretest interviews.
Chapter 3

Methods

Participants

Vosniadou and Brewer (1992) and Sneider and Pulos (1983) found that the widest range of models of the shape of the Earth is found at the ages of 9-10 and that many students have accepted the spherical Earth model by the ages of 12-13. The present experiment required that many of the students have alternative models so that the persuasiveness of the text could be measured. However, if the participants were too young they would not be ready to accept the idea of the spherical Earth due to the conceptual leaps that are needed to solve the inconsistencies between theory and common observation. Therefore, I chose to use participants in third and fourth grade.

There were 130 students enrolled in the study: 31 3rd grade students ranging in age from 8 years 5 months to 9 years 5 months (average age 9 years 2 months) with 12 boys and 19 girls, and 99 4th grade student ranging in age from 9 years 5 months to 10 years 9 months (average age is 10 years 0 months) with 58 boys and 72 girls. The students were in 8 classrooms in four different schools in a medium sized Midwestern city. The demographic information for the district and participating schools are presented in Table 1.

There were two participating fourth grade classrooms in school one, school three and school four. There was one participating third grade classroom in school two and one third grade classroom in school one. The classrooms were assigned to either conduct Collaborative Reasoning discussions or to continue their usual language arts curriculum. The classrooms were assigned using a matched pair technique to make the two treatment groups as similar as possible. One fourth grade classroom in schools one, three, and four were randomly assigned to the CR
condition while the other was assigned to be the control. Since it was not possible to have both third grade classrooms in one school, one of the third grade classrooms was randomly assigned to the CR condition and the remaining third grade classroom was assigned to the control condition.

The eight participating teachers consisted of two men and six women. All the teachers had at least three years of teaching experience. They volunteered to participate, and were recruited by the vice superintendent of the district in charge of literacy education.

**Teacher Training**

All the teachers attended a three hour workshop on the Collaborative Reasoning approach to discussion. The teachers were briefly introduced to the theoretical background of CR, the history of its development, and some of the major findings of previous research. They were then taught how to prepare for a discussion, the teacher’s role during a discussion, and how to incorporate CR discussions into classroom curriculum. Teachers were taught the norms of discussion and argumentation and nine different instructional techniques that can be used to support students in using the norms (Appendix A). Short videos segments were used to demonstrate the instructional techniques and a 15 minute video of a teacher and students having a complete CR discussion was used to familiarize teachers with how the entire lesson might unfold. Teachers were asked to compose heterogeneous discussion groups within their classrooms so that each group was a representative cross-section of the class based on reading ability, gender, ethnicity, and talkativeness. Teachers were assigned to their treatment condition after training and student recruitment. Teachers in the control condition were asked not to begin CR discussions until after the experiment had ended.
Teachers in the CR treatment were observed during the discussions, and the researcher gave targeted feedback and advice as needed.

**Measures**

- Structured Interview diagnosing students’ prior conception about the shape of the Earth
- Additional questions concerning their degree of certainty about their views of the shape of the Earth
- The Stanford Achievement Test Series 10th Edition (SAT-10) - reading comprehension subtest
- The Cognitive Abilities Test (CogAt) Figure Analysis Test (to measure spatial reasoning ability)
- The Reading Fluency Sentence Verification Test
- Need for Cognition Survey (modified for children)

**Pretests.** The Cognitive Abilities Test, Need for Cognition Survey, and the reading fluency test were administered to all the classrooms prior to the beginning of any CR discussions. The tests were administered by the researcher and all three tests took approximately 35 minutes of class time. On a separate occasion approximately one week later and before the beginning of any CR discussions, the students were given the Shape of the Earth Interview with confidence questions. These interviews were conducted by the researcher, and two to four assistants in each class. With permission from the parents of the participating students, SAT10 comprehension scores were obtained from the district. These scores had been administered approximately three months prior to the experiment as part of routine district wide assessments.

The SAT10 was intended to serve as a measure of baseline reading comprehension prior to the treatments. The SAT10 is a widely implemented standardized test (Pearson Inc., 2003). While the test is comprehensive, only the reading comprehension subtest was used in this study.

The CogAt is a commercially available test designed to measure students’ non-verbal cognitive abilities (Riverside Publishing, 2001). The figure analysis subtest was used to measure student’s ability to mentally manipulate objects. This test was employed because it was
hypothesized that producing shape of the Earth mental models requires students to mentally manipulate physical shapes; therefore spatial reasoning might be an important element in developing a mental model of the shape of the Earth. The CogAt was administered and scored according to the guidelines of Riverside Publishing.

The Reading Fluency Sentence Verification Test is designed to measure students' general reading fluency to be used as a covariate during analysis of reading times. The Reading Fluency Sentence Verification Test is comprised of 110 sentences (55 true and 55 false sentences) (See Appendix B for a complete list of items) (He, 2007; Wu et al., in review)). Students are instructed to read the sentences and mark them true or false. The sentences are designed to be obviously true or false so that students will have a very high level of accuracy if the sentence was actually read. The number of sentences insures that few, if any, children can finish all the sentences in the allowed time while not being so long as to discourage children. The number of syllables in the correct items completed in five minutes constitutes the reading fluency score.

The Need for Cognition Survey was added in hopes of controlling for any individual difference between students due to a stable tendency to deeply process ideas. The Need for Cognition Survey short version (Cacioppo, Petty, & Kao, 1984) was originally designed for adult participants. Therefore, the language had to be altered from the original to make it easier for children to read and understand (See Appendix C for a complete list of items). Four of the items were deemed to be outside of children’s experience and could not be made appropriate by simplifying the language. These items were removed resulting in a 14 item survey. Since the language and number of questions has been changed, no attempt was made to compare results from this study and previous work using the NFC survey.
**The shape of the Earth interview.** The shape of the Earth interview was first developed by Vosniadou and Brewer (1992) to probe students’ mental models of the Earth. The questions elicit both verbal responses and drawings. The interview includes factual and generative questions. Factual questions such as “What is the shape of the Earth?” or “Show me where China is on your drawing,” could be answered using isolated facts learned during instruction. While these questions are useful in judging what theoretically important facts the child has acquired, generative questions are needed to understand what these facts mean to the child – how they are used in the child’s mental model. The shape of the Earth interview includes generative questions such as “if you walked and walked for many days in a straight line, where would you end up?” These questions are unlikely to have been taught to the child. In fact, it is unlikely that any of the participants have considered these exact questions before. Therefore, these generative questions force children to use their mental models to simulate circumstances and thereby reveal information about the nature of their mental models even if they themselves cannot explicitly express what those models are.

The majority of the interview questions used in this study were asked exactly as written in Vosniadou and Brewer (1992), but several changes were made. Firstly, a question was inserted between the first and second questions in the interview. After asking students “What is the shape of the Earth?” they were asked, “How do you know the Earth is _____” (the word used by the student in question one was inserted for the blank e.g. round, circle, sphere). This question was inserted because it provided information about whether students changed what they considered the source of their knowledge between the pre and post test. We were especially looking to see if students attributed their knowledge at the time of the post test to the target reading. This question was taken from a different version of the shape of the Earth interview.
(Vosnaidou & Brewer, 1987). In the original interview questions three, four and five were “What is above the earth?” “What is below the Earth?” and “What is to the sides of the earth?” However, Vosnaidou and Brewer (1992) did not find these questions were diagnostic of students’ mental models, because all students gave similar answers. Therefore, these questions were omitted in this interview. In the question “Now I want you to show me where Champaign is,” Champaign was replaced with the students’ home town. Two additional questions were added at the end of the interview, “Does anyone live, here, on the bottom of the Earth?” and “Why don’t they fall off?” The second question was only asked if the student said that someone could live on the bottom of the earth. If the student said that they could not they were asked “Why not?” and if they responded that it was too cold or that there was not food, they were told that visitors could bring special equipment and provisions, and then were asked the question again. This set of questions was also taken from an alternate version of the shape of the Earth interview (Vosnaidou and Brewer, 1987). These questions were added, because the target story emphasizes this issue. Adding this question allowed for us to assess growth in this particular aspect of the model. (See Appendix D for interview questions and protocol).

At the end of the interview, students were asked the following four questions to measure the certainty with which the students held their views:

1) When you think about the shape of the Earth, are you very confused, confused, certain, or very certain?
2) When you try to picture the Earth, is it very hard, hard, easy, or very easy?
3) How sure are you that you have the correct idea about the shape of the Earth, very unsure, unsure, sure, or very sure?
4) How many other people do you think agree with you about the shape of the Earth, very few, a few, a lot, or almost all people?

These questions were added to test the hypothesis that students who had scientifically incorrect mental model at the pre test and retained that model in the post test, might have lower
certainty in the posttest after reading the target story. Furthermore, students who had a scientifically accurate mental model initially might increase in confidence after reading the target story. The confidence questions were derived from the attitude certainty questions composed by Petrocelli, Tormala, and Rucker (2007). These questions are designed to measure the two aspects of attitude certainty: attitude clarity and attitude correctness. Petrocelli, Tormala, and Rucker define attitude clarity as “the subjective sense that one knows what one’s attitude is.” This construct is measured by questions one and two in the questions above. Petrocelli, Tormala, and Rucker define attitude correctness as “the subjective sense that one’s attitude is correct and valid.” This construct is measured by questions three and four in the questions above.

The post-test interview included the same questions plus several more that were designed to provide a self-report from the student of conceptual change, and if they felt the target reading was influential in producing any change. The questions added at the time of the post-test questions were as follows:

5) You answered many of these same questions a few weeks ago. Do you think you changed your mind about any of the questions since the last interview?

6) Did reading the story make you think differently about any of the questions I asked?

*Interviewer recruitment, training and procedures*. The interviewers were drawn from community members, graduate students from the University of Illinois, and undergraduate students from Illinois State University. The interviewers received training on how to conduct the interview and how to use the digital recorder. They each received a copy of the interview protocol and interview recording sheet to review before conducting the interviews (See Appendix D). The interviewers (other than the researcher) were blind to the treatment conditions of the students and the purpose of the study. Other than the researcher, they did not receive instruction about what models of the Earth students might have and what answers corresponded
to those models. The interviewers were informed that they should “ask additional questions only to clarify until you feel you understand the student's intentions.” In addition, they were instructed to try gestures and simple questions such as “can you tell me more about that” and only use more specific questions if all of these techniques failed to produce a comprehensible answer.

The uninformed nature of the interviewers and the careful instructions to avoid leading questions led to less probing than is seen in the samples provided in Vosniadou and Brewer (1992). For example, in the following exchange from Vosnaidou and Brewer’s study, the interviewer probes the student for exactly what they mean by living inside the Earth in order to confirm that the student is using a hollow Earth mental model

Interviewer: Can people fall off the end/edge of the earth?

Student: No.

Interviewer: Why wouldn’t they fall off?

Student: Because they are inside the earth.

Interviewer: What do you mean inside?

Student: They don’t fall, they have sidewalks, things down like on the bottom.

Interviewer: Is the earth round like a ball or round like a thick pancake?

Student: Round like a ball.

Interviewer: When you say that they live inside the earth, do you mean they live inside the ball?

Student: Inside the ball. In the middle of it.

(p. 564)
In comparison, the less active style of the interviewers in this study can be seen in the following interchange. This student is also classified as a using a hollow Earth mental model, but the interview does not use the more extensive probing:

Interviewer: Does it have an edge? Can you fall off?

Student: No

Interviewer: And why don’t you think you can fall off?

Student: Because the earth is a circle and the sides are like a wall?

The ignorance of the interviewers and the careful avoidance of leading questions were emphasized to a greater extent in this experiment, because of the pre and post test nature of this design. It was important that students learned as little as possible from the interview process itself. Furthermore, even though the clinical interview technique is less prone to the effects of repeated questioning compared to forced choice formats (Brewer, 2008), we wanted to side on the side of caution to avoid implying to the students that they needed to change their minds simply because of repeated questioning as has been shown in some child eye witness testimony research (Siegal, Waters, & Dinwiddy, 1988).

The confidence questions were given at the end of the interview. The interviewer held up cards with the question and the choices (e.g. very unsure, unsure, sure, very sure) printed in large type, and the student was asked to indicate which word or phrase would best describe themselves. The cards were used to avoid children forgetting or being confused by the choices given.

The interviews were conducted during class time. Students were taken out of class individually and interviewed either in the hallway, cafeteria, in a corner of the classroom, or other relatively quiet location in the school. The interviews lasted between 8-18 minutes and
were recorded on digital recorders. A record sheet was used for each student that listed the interview questions and provided space for interviewer notes which included gestures and other nonverbal communication. Students not being interviewed continued teacher assigned seat work.

**Interview scoring.** Each individual answer given by the student was categorized according to a scoring rubric. The scoring rubric was based on the scoring rubric developed by Vosniadou and Brewer (1992) and subsequently expanded in later revisions of the survey. The scoring rubrics were designed by Vosniadou and Brewer to describe the full range of student answers. However, this set of students had certain answers that might be of theoretical interest but were not captured by the previous categories so several new categories were added to some questions. (see Appendix E for the scoring rubric).

Once all the answers were coded, the pattern of answers given by each student was matched to the expected pattern for a person with the mental models described by Vosniadou and Brewer (1992) (See Appendix E for the scoring rubric). Consistent with Vosnaidou and Brewer, each answer was considered either consistent with a model, an acceptable deviation from the model, or an unacceptable deviation from the model. Vosnaidou and Brewer define an acceptable deviation as an answer which is “in principle inconsistent with the mental model in question,” but, “can nevertheless be explained on the grounds that it represents a semantic error or is ambiguous with respect to its meaning” (pp. 554). Vosnaidou and Brewer allowed no unacceptable deviations and only one acceptable deviation for an interview to be scored as representing a particular mental model. This strict criterion used by Vosniadou and Brewer was needed for the purpose of the paper which was to identify the existence and consistency of mental models of the Earth in children. This dissertation builds on the findings of Vosnaidou and Brewer, but it is focused on the process of how students’ change their mental models rather than
if they exist. For this reason, we used a slightly different interview protocol that tended to have few follow up questions as described above. This led to many students being considered “mixed” using the initial scoring of Vosnaidou and Brewer. However, this would diminish the ability of the current research to measure growth since mixed scores include students who have unacceptable as well as multiple acceptable deviations from the expected pattern. We assert that having acceptable deviations, although not ideal, is still a closer approximation to the model than an unacceptable deviation. Therefore, we created two categories for each mental model: strict and loose. Strict was given to interviews that had one or less acceptable deviations and no unacceptable deviations from the model. Loose was given to interviews that had no unacceptable deviations, but had more than one acceptable deviation. For the loose category, claims of ignorance (e.g. “I don’t know”) and missing data points were considered acceptable deviations, but for strict they were not. An interview was only classified into a loose model, if it did not qualify as any other model. In only three cases, did an interview qualify as both loose sphere and loose hollow due to a number of exceptionally vague answers. These interviews were classified as mixed. All the interviews that did not qualify as loose or strict instances of one of the expected mental models were classified as mixed.

**Spherical Earth model.** This section will summarize what answers were considered consistent with the spherical model of the Earth. The differences between this study and the original scoring rubric found in Vosnaidou and Brewer (1992) will be highlighted. For a list of all the answers for each question see Appendix E.

Question One “What is the shape of the Earth?” – The answers round, sphere, and oval were all consistent with the model. Circle was marked as an acceptable deviation because students might use this word to refer to either a flat disc or a sphere.
Question Two “How do you know the Earth is round?” – This question was not used for either model fit or for comprehensiveness.

Question Three “Which direction do we look to see the earth?” – The answers down, sideways, everywhere/all around and in the back/forward/straight ahead were all consistent with the sphere model. Up was considered an acceptable deviation. In addition, many students in this sample gave the answer “north.” This was considered a synonym for up and was also considered an acceptable deviation.

Question Four “Can you draw me a picture of the Earth?” - Circles or oval drawings were considered consistent with the model. All other types of drawings would have been considered unacceptable deviations, but all the participants drew either circles or ovals.

Question Five and Six “Now on this drawing show me where the moon and stars go. Now show me where the sky is on your picture.” – Our protocol differed from the original in that we often prompted children to draw clouds where they thought the sky should be. For this reason several different variations appeared in the data for this experiment that were not given in the original scoring rubric (Vosniadou & Brewer, 1992). However, the same principals apply that all drawings in which the stars appeared above or around the earth and in which the sky appeared on, around or above the Earth were considered consistent with the spherical Earth model. Drawings in which the stars and/or moon appeared on the earth or in which the sky appeared only below the Earth were considered unacceptable deviations. Figure 4 shows examples of participants who qualified as having spherical models and had drawings of the most common patterns of one, three, five, six, and seven (see Appendix E for illustrations of each answer).

Question Seven “Show me where the people live.” – All the drawings in which people were inside the circle or standing on the perimeter were consistent with the sphere model.
Drawings in which people stood on a flat line inside or outside the Earth were unacceptable deviations. However, there were only three students who did not respond that people live inside the circle.

Question 8 “Here is a picture of a house. This house is on the Earth, isn’t it? How come the Earth is flat here, but before you said it was round?” – We added several new categories for this question because we wanted a more detailed account of the range of answers. Students who gave an answer that showed that they understood the conflict between a round Earth and flat looking ground (answers one and four) were considered consistent with the model. Saying that the Earth is round, but that we live on flat pieces, and students who said that the Earth looks flat on the outside, but is really round on the inside were considered acceptable deviations. Students who responded that the Earth looks flat so that we don’t fall off were also considered acceptable deviations. All other answers were unacceptable deviations. Here are two examples of answers to question number eight from students that were categorized as having a spherical model of the Earth. The first had an answer the strictly fit the model rubric, and the second student had an answer that was an acceptable deviation.

Example of an answer to question eight that was coded as consistent with a sphere model:

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?

Student: House. Yeah.

Interviewer: How come the earth is flat here, but before you said it was a sphere?

Student: Because, you could call that picture incorrect. Because the earth seems flat, you know? Unless you are in a mountain, hill or something. But it seems flat because the earth is so big compare to this one human, and you can’t tell it’s bending.
Example of an answer to question eight that was coded as an acceptable deviation from a sphere model:

Interviewer: This is a house, and the house is on -

Student: The earth.

Interviewer: A flat earth, right?

Student: Yes.

Interviewer: How come the earth is flat here, but before you said it was a sphere?

Student: Because the earth is different land forms and that’s probably flat instead of a hill like that. And some parts of the earth are curved and bumpy and some parts are straighter than others.

Question Nine “If you walked and walked for many days in a straight line, where would you end up?” – Unlike the original scoring, we gave this an individual score from the answers to questions ten and eleven. We considered the answer that you get back to where you started or that you go in circles as consistent with the sphere model. We considered the answer that you just keep going or that you end up in a specific location as an acceptable deviations. All other answers were unacceptable deviations.

Questions 10 “Does the Earth have an end or edge? Could someone fall off the end or edge of the Earth? Why? Where would they fall?” – In the original scoring these were considered separately, but in our scoring they were considered together. Answers that the Earth did not have an edge because it was round, because the Earth is moving, because you go around to where you started, or because if there was you would fall off, are all considered consistent with the model. Acceptable deviations include the answer that there is an edge, but you cannot fall of because of gravity. A student who stated that there was no end but could not give an explanation was
considered an acceptable deviation. All other models were unacceptable deviations. Below are examples of the common answers one, two, and six.

Example of answer one:

Interviewer: Does the Earth have an end or edge?

Student: No

Interviewer: Why not?

Student: Because it’s a sphere and spheres don’t have edges.

Example of answer two:

Interviewer: Does the Earth have an end or edge?

Student: No.

Interviewer: Why not?

Student: Because it’s three dimensional like a ball, you can run a car all around, all the way around it, and still wouldn’t just… it’s not like a piece of paper, say this was a car, to, to, to, oops I run out of land so I’m going to fall off in the space or something, you can’t fall off that’s not… Unless you are talking about mountains, mountains are point up, that’s a cone.

Example of answer number 6 which is an acceptable deviation. Despite the fact this person says that there is an edge to the Earth, they demonstrate understanding of the spherical model:

Interviewer: Does the Earth have an end or edge?

Student: Well, it keeps going around and around, but I guess you could consider that, like, an end or an edge.

Interviewer: Okay.
Student: ‘Cause, it doesn’t go on forever. The galaxy does, but the Earth doesn’t. It’s just a big big big big big big big millions of people fitting big big big big thing of Earth. Which is called Earth.

Interviewer: Ah. Could someone fall off the end or edge of the Earth?

Student: No.

Interviewer: So I think you’re saying like the outside of the Earth is kind of like the edge of the Earth. Is that what I’m hearing?

Student: Well, sort of, but you can’t fall off it. ‘Cause gravity is holding us down.

Question Eleven “(a) Now I want you to show me where your home town is on your picture. (b) Now I want you to show me where China is on your picture.” – Drawings that showed both cities to be inside the circle, or drawing that showed one city inside the circle, but the other location was unknown, were considered consistent with the sphere model. Answers that placed either city outside the circle were considered unacceptable deviations. All other drawings were unacceptable deviations.

Question Twelve “Now tell me what is down here below the Earth.” – Sky, celestial objects, and/or space were considered consistent with a sphere model. All other answers were unacceptable deviations.

Question Thirteen “Does anyone live, here, on the bottom of the Earth? Why don’t they fall off?” – Answers indicating that people can live on the bottom but would not fall off because of gravity, or who did not give a reason were considered consistent with the sphere model. Answers indicating that people can live on the bottom and would not fall off, because the Earth is round were considered acceptable deviations. Answers indicating that people cannot live on the bottom, but if they could they would not fall off because of gravity were also considered
acceptable deviations. All other answers were considered unacceptable deviations. Here is an example of an answer from a student who was categorized as having a spherical model. They gave a gravity answer to question thirteen:

Interviewer: Does anyone live, here, on the bottom of the Earth?

Student: Yeah, that is Antarctica

Interviewer: So, do people live in Antarctica

Student: Well that’s kind of more down, I just don’t have the room to make such a big picture that reaches South America down there…

Interviewer: Why don’t they fall off?

Student: Because gravity, ahm – I can’t say that it actually pulls things to the core, because that’s not what happens, but it does, gravity holds things on to the crust of the earth

Example of a student who was classified as having a spherical model of the Earth. This student gave answer two for question thirteen. They required some prompting by the interviewer to explain their thoughts:

Interviewer: And how about this here, what’s this?

Student: That’s Antarctica.

Interviewer: And does anybody live here? How can they live there without falling off?

Student: They’re not—it’s still flat. And it’s not upside down. See how that is?

Interviewer: It’s not upside down?

Student: No. It looks like it’s not, but it’s not [Pause]

Interviewer: If we were in Antarctica, would we feel like we were standing on our head?

Student: No.
Interviewer: Why not?

Student: ‘Cause it don’t feel like you’re upside down, it’s flat. Down here it’s flat, but it’s not space.

Interviewer: And you don’t—why don’t you fall off?

Student: ‘Cause it never ends.

Question Fourteen “Is there anything more you would like to say about the Earth?” – This did not get a separate score, but all answers given here that were relevant to other questions were counted for that question.

**Hollow Earth model.** This section will summarize what answers were considered consistent with the hollow model of the Earth.

Question One “What is the shape of the Earth?” – In addition to round, sphere, and oval, circle was considered consistent with this model, because this is what the ground would look like from above a hollow Earth.

Question Two “How do you know the Earth is round?” – This question was not used for either model fit or for comprehensiveness.

Question Three “Which direction do we look to see the earth?” – In addition to down, sideways, everywhere/all around, and in the back/forward/straight ahead, up and north were also considered consistent with this model, because people inside a hollow Earth can look up to see the top hemisphere of the Earth. All other models were inconsistent with the model.

Question Four “Can you draw me a picture of the Earth?” – Circles or oval drawings were considered consistent with the model. All other types of drawings would have been considered unacceptable deviations, but all the subjects drew the Earth as a circle or an oval.
Question Five and Six “Now on this drawing show me where the moon and stars go. Now show me where the sky is on your picture.” – All the drawings accepted for the sphere model where considered consistent with the hollow model, and in addition, the drawing with the stars, moon and sky inside and at the top of the Earth were also considered consistent because hollow Earth models allow for the top hemisphere to contain the sky. Figure 5 shows an example of the only type of drawing that was considered consistent with a hollow Earth model but not a spherical model.

Question Seven “Show me where the people live.” – Drawings in which the people were inside the circle, or on a line inside the circle were considered consistent with the model. All other drawings were unacceptable deviations.

Question Eight “Here is a picture of a house. This house is on the Earth, isn’t it? How come the Earth is flat here, but before you said it was round?” – Question eight proved to be particularly diagnostic of hollow Earth models. Children who explained that the Earth is round on the outside but flat on the inside were considered consistent with this model. The answer that the Earth is round, but that we live on a flat part was considered an acceptable deviation, because this could be a poorly articulated explanation of the hollow model. A few additional answers were considered acceptable deviations because they were consistent with an Earth that has an inside and an outside even though they did not explicitly mention this. Other answers were considered unacceptable deviations. Here is an example of an answer for question eight that exemplifies the hollow Earth model:

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?

Student: Mm-hmm..

Interviewer: How come the earth is flat here, but before you said it was a circle?
Student: Because, this is the sidewalk probably, and it probably has some grass and stuff.

Interviewer: Okay, but look at this whole hallway, it’s all flat, how could there be… and when I drive I see all these fields, all these farms…

Student: Because we’re in the earth.

Interviewer: We’re in the earth?

Student: Yeah.

Interviewer: So tell me about that. So how come it looks flat when we’re in the earth?

Student: Because if it was flat, you don’t have like sticking shoes to go around, if it was circle what if you wanted to go like under, we would have to walk upside down. [Interviewer: Upside Down!] And then everybody would fall.

Question Nine “If you walked and walked for many days in a straight line, where would you end up?” – For the following two questions, we depart slightly from the scoring rubric of Vosnaidou and Brewer (1992). They described the hollow Earth model as either not having an end/edge or that you cannot get there or can only get there by going up through the top. We also believe that a hollow Earth model could include a permeable upper hemisphere made of sky. For this reason we made the answer that a person could not get to the end or edge as the most model consistent, but we also counted as acceptable deviations answers stating that a person would fall off the Earth, keep going, or reach a specific spot. All other answers including that a person would return to where they started were unacceptable deviations.

Questions Ten “Does the Earth have an end or edge? Could someone fall off the end or edge of the Earth? Why? Where would they fall?” – Continuing from the position taken on question nine, a student can think that either it does or it does not have an end or edge and still
get credit for this question. If the student says that there is an edge, they would only be considered model consistent if they say that a person cannot get to the edge because it is too high up to reach or that there is a barrier. If the student thinks there is an edge, and they believe that a person could fall off the edge into space, than that is considered an acceptable deviation. If a student states that there is no end or edge because the Earth is round, this is also considered model consistent. All other answers were not considered model consistent. Here is an example of a student whose model is categorized as hollow and expressed the belief that people cannot reach the edge of the Earth, because there is a wall:

Interviewer: Does the Earth have an end or edge?
Student: mm
Interviewer: No? Why not?
Student: Because we are, you need to walk like ten years for go on the side. And then, wait, what was the question?

Interviewer: Does it have an edge? Can you fall off?
Student: No

Interviewer: And why don’t you think you can fall off?
Student: Because the earth is a circle and the sides are like a wall?

Question Eleven “(a) Now I want you to show me where your hometown is on your picture. (b) Now I want you to show me where China is on your picture.” – Drawings that placed Bloomington and China inside the circle were considered consistent with the model. Drawings that placed either location outside the Earth or placed one location on the invisible back side of the Earth were considered unacceptable deviations.
Question Twelve “Now tell me what is down here below the Earth.” – As with the sphere model, sky, space and/or celestial objects were all considered consistent with the model. However, in the hollow Earth model, dirt and water were considered acceptable deviations. All other answers were unacceptable deviations.

Question Thirteen “Does anyone live, here, on the bottom of the Earth? Why don’t they fall off?” – The confusion here is what the student considers as the bottom. However, if a student stated that someone could live on the bottom, they needed to explain that they would not fall, because they were inside the Earth in order to be model consistent. Students who said they would not fall because of gravity, the roundness of the Earth, or because the Earth was flat, or because they were inside houses, were all considered acceptable deviations, because they could all be interpreted as being part of a hollow mental model of the Earth. If the student stated that they could not live on the bottom, because they would fall off, or because they were upside down, or because they would be underground, were all considered model consistent. Other answers were not consistent with the model.

Question Fourteen “Is there anything more you would like to say about the Earth?” – This did not get a separate score, but all answers given here that were relevant to other questions were counted for that question.

Only six of the interviews representing four individuals had models which were not classified as either sphere, hollow or mixed. Those individuals are described next.

*Flattened Earth model.* The student who was categorized as having a flattened Earth model had very similar answers to a hollow Earth model. This student drew a picture with a line across the picture which might have indicated a hollow Earth or a sky near the top of the Earth as
in a flattened model. The deciding factor was the following statement from this student’s answer to number eight:

Student: Oh yeah, the sky [pause] well [pause]. Maybe if we were somewhere like on the top maybe it’s flat cause, because, it’s, if you would be on it it would be flat.

Certainly, this student is not an exemplary example of the flattened Earth model, but it is the most likely fit.

**Dual Earth models.** Of the two interviews that were categorized as dual-earth models, only one drew two earths (Figure 6). The other drawing was similar to the drawings of the disc shaped Earth models, but this student’s other answers differentiated her model. In particular, the answers to number eight and thirteen were diagnostic. For number eight, she indicated that there were two places space and land:

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?

Student: mhm

Interviewer: How come the earth is flat here, but before you said it is circle?

Student: Because this is out of space, and space not everything gets flat and in space is different from land because land is flat. Some lands are kind of hills but the earth is flat

Similarly for question 13 she suggests that there are two places both called Earth:

Interviewer: Does anyone live, here, on the bottom of the Earth? Now you said nobody lives here, but do people live, let’s say: here or here?

Student: Yes, cause like people live in south and north and stuff.

Interviewer: How can they walk there for few days without falling off?!
Student: Because they are not, like in outer space they are in a country, they are not like they are living on the actual earth, they are living on earth but they are living on homes.

**Disc Earth models.** Students who were categorized as having a disc shaped model made drawings in which the sky, stars and moon were drawn inside the circle representing the Earth, or in which only the sky was above the Earth. (Figure 7)

Students with disc shaped models gave answers to question eight that were very diagnostic. Here is an example:

Interviewer: How come the earth is flat here, but before you said it was a circle?

Student: Yeah, it’s a circle and it’s, like, flat, well, on a map it’s flat, and it’s kind of spread out a little.

Students with flat earth models often gave answers to question number nine that indicated that a person who walked and walked would endlessly go to new locations. However, for question ten, they indicated that people could fall off the Earth although they differed in where a person would go if they fell. The following example is from a student who insisted that there was no end to the Earth, because it is infinitely large, but inferred that if there was an end a person would fall off:

Interviewer: Okay. And, if you walked and walked for many days in a straight line, where would you end up?

Student: In a different continent than where you started. Like if you started in North America, you might stop in South America.

Interviewer: Okay, keep walkin’. Keep walkin’ and walkin’ for days and days—where would you end up?
Student: Australia?

Interviewer: Okay, keep walkin’.

Student: [Pause] You’d just keep going and going and going and going. And you wouldn’t stop.

Interviewer: Okay. Why?

Student: Because the Earth never stops.

Interviewer: Okay. Tell me more about that.

Student: The Earth never stops because if the Earth stopped, we wouldn’t know where to go. Because it’d just be air around us.

Interviewer: Okay. So, is there an edge to the Earth, or an end?

Student: No.

Interviewer: No? So, why not?

Student: Because the Earth is everywhere, and if you stopped at the end of the Earth, you wouldn’t be anywhere, there would be no ground or anything.

Students with disc models had no problem with people living on the bottom of the Earth. When asked why people do not fall off the bottom, they answered in a way that would be true of any location on the Earth:

Interviewer: Does anyone live, here, on the bottom of the Earth?

Student: Yeah. No.

Interviewer: Why not?

Student: Because it’s water.

Interviewer: What if they had a boat?

Student: Then they could ride in the water on the boat.
Interviewer: Why don’t they fall off?

Student: Because it’s the water, and water don’t go down, unless it dries up.

Rectangular Model

No interviews were consistent with the rectangular model described in Vosnaidou and Brewer (1992). For this reason the scoring criteria for these models will be omitted here.

**Mixed models.** As stated before, the mixed model designation was designed to capture students who gave some answers that belonged to one model while other answers belonged to another model. In addition, since my analysis allowed for loose fit models, the mixed designation was also reserved for models that loosely fit the rubric for more than one model. However, this later circumstance only occurred three times.

One student who exemplified the mixed model on his pre-test interview, demonstrated a general understanding of the spherical shape of the earth, but showed that she had not yet come to terms with violations of the two of the fundamental assumptions identified by Brewer (2008) as interfering with adopting the spherical model of the Earth: things that appear flat are flat and gravity pulls things in a universal downward direction. Holding to a spherical Earth in principle without inhibiting these assumptions resulted in a very incoherent model of the Earth. On the one hand, she easily concluded that a person walking in a straight line would return to where they started, that the Earth has no edge, and that China is on the other side of the Earth from Illinois. At the same time, she stated that the surface of the Earth looks flat in a picture because the grass makes it flat. Furthermore, she gave an explanation for why people do not fall off the bottom of the Earth that seems to suggest a hollow Earth:

Interviewer: Does anyone live, here, on the bottom of the Earth?

Student: No because it’s too cold down there.
Interviewer: It’s too cold? What about like right here, or…

Student: Maybe. No.

Interviewer: Ok, no, if they live here, why wouldn’t fall off?

Student: Because, the Earth from outer space, there is like, there is all like water down here. But if you go down into the Earth, then you will not be able to fall off over, like out of the Earth.

Interviewer: Can you tell me a little bit more about that?

Student: If – like people live on the earth but they don’t live on the outer space, but they live down and they were like, you can’t breathe and stuff because in outer space you can’t breathe so you will have to either be an astronaut to go out of the Earth

Interviewer: Ok, can’t live, you don’t fall off, because you are inside.

Student: mhm

This student’s statement that you need to “go down into the Earth” to not fall off the bottom suggests that she is using some version of the hollow Earth model to allow her to resolve the conflict between a spherical Earth and people living on the bottom. However, the many starts and stops and repetitions might also indicate that this student already felt uncomfortable and uncertain about her answers.

Another student who had a mixed model during the post-test, illustrates how a mixed model could result from the incomplete process of incorporating the information from the story with their previous beliefs. This student had several answers which showed contradictions even while answering the same question. Her answer to the third question is an example of this:

Interviewer: Which direction do we look to see the earth?
Student: Well, we’re on it, so you can’t really see it.

Interviewer: Well, if you walked outside, would you look up or down.

Student: Up.

Interviewer: Why Up?

Student: Because that’s where all the stars and the planets are.

Interviewer: If you walked outside which direction would you look to see the Earth? Up or Down?

Student: Well if you looked down you could see the… down then, because the ground is the Earth, because it says that, in the story, you stand on the Earth.

You’re not inside of the earth your on top of the Earth.

Interviewer: So you’re on the Earth.

Student: Yeah, so down.

Despite the somewhat leading nature of the interviewer’s follow-up question “up or down?” The student seems to be thinking out his answer on the spot. He seems to be reflecting on what he learned during the story and comparing this to their previous beliefs about a hollow earth. It is telling that the story does not actually mention that we do not live inside the Earth. Despite this realization this same student has great difficulty with question nine:

Interviewer: If you walked and walked for many days in a straight line, where would you end up?

Student: Possibly right here at the end of the earth?

Interviewer: What happens when you get to the end of the earth then?

Student: You can’t go any farther.

Interviewer: Because why?
Student: Because it ends.

Interviewer: What does the end of the earth look like?

Student: I’ve never really gone to the end of the earth so I don’t know.

Interviewer: Would you fall off?

Student: No.

Interviewer Why not?

Student: Because it’s not like... because... [pause] I don’t really know.

Once again, while this student seems to be reverting to their hollow Earth pre-test beliefs, they’re confusion and uncertainty illustrates their struggle. This student is most likely categorized as mixed, because he is in the process of reflecting on the issues.

Coding for model change. The coding of the interviews was conducted blind to whether each interview was administered before or after the treatments (although some interviews revealed that they were post-tests by referring to the story). After coding was completed, the results were matched to the students’ identity to investigate how students’ models of the Earth changed. Examples of model change are given in the Results section.

The Discussion Treatments

The four classrooms in the regular instruction condition were asked to continue their usual language arts curriculum for the two weeks between the pre- and post-tests. Each of the teachers in the regular instruction condition was asked to schedule a lesson in which they had a discussion. The lesson was to be at least 20 minutes long. Pains were taken to explain that “discussion” could take any meaning that they chose and that what I was looking for was a sample of their typical language arts instruction. These discussions were videotaped so that the nature of the baseline discussions in the control group could be described.
The four classrooms in the CR discussion treatment conducted two CR discussions each week for two weeks. All the classrooms discussed the following stories in the same order as determined by the researcher: “Trip to the Zoo” (Reznitskaya & Clark, 2001), “Crystal’s Vote” (Nguyen-Jahiel, 1996), “A Coat for Mr. Snowman” (Miller, 2009a), and “Deep Water” (Miller, 2009b). The first two stories have been used in previous CR studies and have proven to produce dynamic discussions in many classrooms (Wu et al., in press). They were chosen to allow students to have a successful introduction to the CR format. Trip to the Zoo explores the ethical dilemma “Are zoos good places for animals?” Crystal’s Vote explores a policy issue in a school which must choose to either purchase new textbooks or computers.

The third and fourth stories were written for this experiment, and they were designed to be similar to the target story so as to make the social context of CR as salient as possible when the children read the target story. These stories were written to have the same format as the target text in which children disagree about a scientific issue and its practical implications. The stories were also similar to the target story in that multiple sides of the science issue were presented, the proponents of each side presented reasons backed with evidence (some of which were based on pictures), and a conclusion or consensus was never reached in the story. While the format we designed to be similar, the topics chosen were different enough from the subject of the shape of the Earth as to make any transfer of content very unlikely. The two dilemmas chosen were based on concept cartoons presented by Keogh and Naylor (1999).

The story “A Coat for Mr. Snowman” (Miller, 2009a) (See Appendix F) is about three children that are building a snowman. One child suggests putting a coat on the snowman so that it will not melt, but another child complains that it will melt the snowman because coats are warm. The third child thinks that coats only work for people. The students are asked “Should the
children put a coat on Mr. Snowman?” The story “Deep Water” (Miller, 2009b) (See Appendix G) is about three boys who are fishing. One child wants to go from the shallow end to the deep end so that the boat will float higher, but another child, John, disagrees and worries that the boat will sink in deep water. A third child thinks that the depth of the water makes no difference to how a boat floats. The students are asked, “Is John right, will the boat sink if it is the deeper water?” Both stories were pilot tested to insure that students of this age range would find the issues controversial and engaging (See Appendix F and G for the text of these stories).

**Target Reading, Comprehension Test, and Self-Assessment of Depth of Processing**

A week after the final CR discussion, all the students read the target text and took the immediate post-tests. Students were called out in groups of 3-5 to the hallway, cafeteria, quiet corner of the room, or another quiet location in the school. Each child sat in front of a laptop computer. The instructions, practice story, target story, comprehension test, and self-assessment were presented using the software, SuperLab Version 4. This program automatically records the key strokes of students and the times between key strokes. An attendant was present to answer questions or to diagnose software or hardware problems that arose.

When students began they were instructed on how to press the space bar to advance through the reading. They then were familiarized with the procedure by reading a practice story, “The Tortoise and the Hare” (adapted from Jacobs, 1902), using the same clause-by-clause reading method used for the target story. In this experiment, each clause appears centered on the screen. All the clauses are presented in one line so that students do not need to shift their eyes to begin a new line. When the student presses the space bar the previous clause is replaced by the new clause. Students cannot go backwards in the text. Occasionally a picture appears after a clause. The student presses the space bar and the next clause appears following the picture. Once
the students finished the practice text, the computer instructed the student to ask the attendant any questions that they had about the instructions. When they were ready, the students proceeded to the target story.

The target story is written in an argumentative style in which two girls talk about the shape of the Earth and give arguments for a round and a flat model of the Earth. The passage consists of 569 words, 112 clauses, and 6 pictures. The passage has a Flesch-Kincaid Grade level of 4.3. See Appendix I for the passage.

The story was divided into clauses which are defined as a group of words that expresses a proposition, and contains a subject and a predicate. However, some clauses were combined when dividing a sentence might cause confusion or when a single clause did not contain any pertinent information to the topic. This was particularly important for the clauses describing who spoke such as “Jill said” or “Maria asked.” These clauses were always written directly before the quotation and were always included with the clause that followed. If these clauses were presented individually, then students would need to remember who was speaking when they proceeded to the next clause. This would be an unnecessary cognitive task that would interfere with measuring reading times.

Interspersed within the story were six pictures. Each picture was preceded by a sentence that explained what was in the picture, but it did not explicitly indicate that a picture would appear. When the student pressed the space bar after the description, the pictures appeared alone on the screen. The student then needed to press the space bar again to advance to the next clause after the picture. The amount of time students watched each pictures was recorded in the same way as the clauses.
Sentence Verification Test. After reading the passage, students completed two sets of questions on the computer. The first set was designed to measure the students’ comprehension of the shape of the Earth reading passage. The test uses a version of the Sentence Verification Test (SVT) format (Royer, Carlo, Dufresne, & Mestre, 1996). Royer et al. adapted a technique developed by psychologists for the study of discourse processing (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Schmalhofer & Glavanov, 1986) to the study of students reading of science texts. The SVT test consists of four types of items: original, paraphrase, meaning change and distractor. Origin sentences are exactly as written in the original text. Paraphrase sentences retain the meaning of the original sentence but are written using different words. Meaning change sentences are similar to the original sentences, but parts of the sentence have been changed so that the meaning of the sentence is not the same as the original. A distractor sentence has a similar style to sentences in the story and has the same theme as the story, but has a meaning unrelated to any actual sentence in the story. This combination of items is designed to discriminate between participants who might have a memory of the syntactic features of the text but have not retained a gist of the meaning of the text. Without an accurate gist, participants will be likely to inaccurately mark paraphrases as false and meaning change sentences as true.

The SVT for this story consists of 16 sentences (four of each type). Students were asked to mark “TRUE” if the sentence has the same meaning as something Jill and Maria said in the story, or to mark “FALSE” if it has a different meaning from what Jill and Maria actually said in the story.

Self-Assessment Measure. After the comprehension test, students completed a self-assessment of depth of processing. These questions are intended to measure the participants’ own subjective experience of how emotionally and metacognitively engaged they were during the
reading (see results for questions). The survey is composed of 12 Likert scale items (5 point scale). Each item is a sentence describing an approach to the reading passage (i.e. “I put a lot of effort into reading and understanding this story.”). Students were asked to rate how true the statement is for them on a 5 point scale from “not at all true” to “very true.” These items were drawn from the Metacognitive Awareness of Reading Strategies Inventory (Mokhtari, & Reichard, 2002) and the Motivated Strategies for Learning Questionnaire (Pintrich, & De Groot, 1990). However the items have been modified to serve the purposes of this specific experiment. For example, only items that would apply to a single reading passage were used rather than items that addressed student’s overall study and reading habits.

**Posttest Interviews.** Students in each class were randomly assigned into two groups: an immediate posttest or delayed posttest. The immediate and delayed comparison was included to test the prediction of the ELM and the CKRM that students who have had deeper processing of the lesson will have stronger and more persistent conceptual change. The delayed posttest was included to test for persistence. A between subjects design is used, because a within subject test would require each participant receive the interview three times. Despite the open ended format, this additional repetition would greatly increase the chances of children changing their answers in an attempt to please the interviewer.

Immediately following the intervention, one half of the students retook the shape of the Earth interview. The other half of the class was given an epistemological interview. The epistemological interview was developed by Mansfield and Clinchy (2002). The epistemological interview served as a placebo interview. In this way, students not getting the immediate shape of the Earth posttest would be less likely to anticipate the interview in the future, because they were
also being called out to be interviewed like the other students. The epistemological interview will not be analyzed for this dissertation.

The posttest shape of the Earth interview was identical to the pretest interview except for the additional questions added to the confidence questions (see above). In addition, the interviewer was instructed to say the following before starting:

A few weeks ago you answered some questions about the Earth. I am going to ask you some of those questions again, to find out what you think about the Earth now. Just like last time, you are also going to make a drawing, and I am going to tape record us so that I can remember what you said. Do you have any questions? Are you ready to start?

This paragraph was intended to acknowledge that the questions were the same without telling the participant any additional information about why they were being asked the questions. To further limit the awkwardness of a repeated interview and to limit any effect of repeated questioning, a different interviewer was used for each student for the pre and the post interview. In this way, no one was interviewed by the same person twice.

**Summary of Procedures and Timeline**

Table 2 gives a summary of the procedures and their order. It also lays out the amount of time used for the experiment which was 11 hours over 6 weeks for classrooms doing CR discussions and 5 hours over 4 weeks for classrooms doing the regular curriculum.
Chapter 4

Results

The following section will present the analysis and results for each of the measures. I will start by describing the results of the pretest measures and any differences that exist between condition groups on these measures. I will then give a description with examples of the way that the discussion approaches unfolded in the classrooms. Finally I will describe the analysis and results of the post test measures in the order that they administered: reading time, SVT comprehension, self-assessment survey, and shape of the earth interview.

Pre-test Measures

The pre-test measures were designed to determine whether the groups were comparable before the intervention and to serve as covariates when analyzing the four posttest measures (reading time, comprehension, self-assessment, and shape of the Earth interview). Six of the students did not have Stanford Achievement Test scores, because they arrived in the district at the beginning of that year. These students' SAT 10 scores were imputed using a regression equation based on the other three pre-tests and age and gender.

The means reported by discussion and announcement conditions are listed in Table 3. The only significant difference between means was that the SAT-10 was significantly lower for students in regular instruction classrooms than students in CR classrooms (t=-2.26, p=0.026). Although random assignment was attempted, the regular instruction classrooms had more students for whom English was a second language. It is hoped that using SAT-10 as a covariate will partially control for this flaw in classroom assignment.
CR Discussions

The CR discussions varied among classes in both style and quality, but overall the teachers had a high level of implementation. All of the teachers were prepared for discussions and reflected the norms and procedures given to them during training. It is beyond the scope of this dissertation to give a thorough analysis of the discussions, but one discussion is offered as an illustration (See Appendix H for full transcript). This discussion is about the story “A Coat for Mr. Snowman.” At the beginning of the discussion, the teacher gave a short list of norms that she particularly wanted to concentrate on during this discussion which included inviting quiet students to talk, listening to others reasons carefully, and staying on topic. The first students to speak all stated that the children should not put a coat on the snowman. One student thinks that coats make things warmer and therefore would melt the snowman. Another student, Gail, thinks that coats keep out the cold, and that will make the snowman melt. After several minutes of exploring these views Sam states that he does want them to put the coat on the snowman because coats just keep in the warmth that people produce:

Gail: And what I put is you shouldn’t put a coat on a snowman because the coat is made to block out the cold. And they want the snowman to stay cold so it won’t melt. It just doesn’t make any sense unless you want it melt.

Sam: It already is cold. It already is cold. If they put the coat. I want – My answer is yes they should.

Barbara: Even if it is cold the sun will heat up the coat.

Sam: It blocks. The whole point of the coat is to block the heat out.

Barbara: Yeah but if you wear a coat in the summer it’s not going to keep the cold in.
Sam: That’s because you aren’t cold. We’re warm blooded. We adjust to the climate.

Many of the episodes in the remainder of the discussion consist of thought experiments in which students propose situations and then different students argue about how the situations would work out given their different positions. Some of these episodes are aborted, because they spiral off topic as students discuss the details of the hypothetical situations. However, with the help of the teacher some of these thought experiments continue long enough for important issues to be revealed. In the following episode Amil describes a situation that allows the group, and particularly Gail, to work through the question of if coats make their own heat.

Amil: Wait, I think that they shouldn’t put a coat on the snowman because what I’ve done in the past when I was a little kid I used to like shoving big balls like snow, I used to put snow on my coat to see what would happen and it just melts in my coat.

Sam: Because you’re body heats up.

Gail: That is true.

Amil: I know, but even though. Because one time I accidently got my coat in the snow. Like a lot of snow came into my coat, and it just melted.

Sam: Was the coat zipped up?

Amil: No.

Sam: If it went into the sleeves because the sleeves are warmer than just the snow.

Amil: It went into the white part.

Gail: Well if it melted because if you wore it and then it still had some heat that you gave onto it and then you put the snow on it, it could still melt.
The discussion continues and the teacher asks students to reflect more on the text, but throughout the discussion the students returned to this question of Amil’s coat and why it melted the snow. Near the end of the discussion they visit the question one last time. At this point, Barbara is as convinced that the coat makes its own heat as she was at the beginning, but Amil and Gail have been swayed by the discussion of Amil’s coat. Amil struggles to articulate his ideas, but they spur Gail to an epiphany:

Amil: Well, I sort of agree with Sam a little bit now, because if you really think of it, the coat gets wet because of the snow so that and the snow’s cold water so when you put the cold water on the snowman it might have a good chance of –

Sam: It just makes more snow.

Amil: What?

Sam: It has more snow because –

Barbara: It’s not all wet, it’s dry and it’s warm because it is sitting in a closet and you put it on the snowman and it makes it warm.

Amil: I know but to like as it makes it warm the water will go into the coat right? So then the coat will get cold and then like there –

Barbara: No it melts on the coat.

Gail: It melts on the coat and makes the coat wet and –

Barbara: Then it might drip out then

Amil: Yeah but there is some water.

Gail: And the coat absorbs.
Sam: Actually it keeps it cold because if it is already cold it just keeps the heat out of it. And it keeps the cold.

Amil: Yeah, it takes a little bit, it takes a little bit of time, a few minutes to cool down because –

Gail: I see.

Teacher: What do you see that he means?

Gail: Well, if you like put a snowman with a coat on it the coat on the snowman, and the snowman doesn’t have any like natural heat inside it so it won’t like make it warm, and the coat warm and just make - depending on the weather, whether if it’s cold or not. If it’s cold and it will just stay, but if you put it on a human, a warm-blooded person, the heat will spread to the coat which will make the coat warm and will in turn make you warm too.

While most students did not have the articulated conceptual shift that Gail did in this discussion, even students that retained their beliefs like Barbara learned about the alternative issues inherent in the issue. Furthermore, the students learned how to maintain a type of scientific argumentative discourse in which participants identify evidence and then challenge each other over alternative explanations for the evidence.

For the purposes of this experiment, this discussion is a good example of the increased level of social accountability created by CR discussions. Students had to speak for long stretches. They were expected to have an opinion and to be able to express that opinion with reasons. They were challenged by both the teacher and repeatedly by fellow students. This is very different from the regular classroom discussions which will be described next.
Regular Classroom Discussions

The discussions in the regular instruction classrooms took on a variety of forms, but all the discussions reflected the strong district wide literacy program. The literacy program was aimed at teaching students effective reading comprehension strategies such as making predictions, asking questions, making inferences and making connections between the text and their prior knowledge from other texts and from personal experiences.

Two of the classrooms had very similar discussions. In both classrooms, the teacher met with small groups of three students while the other students did independent seat work. In both classes the teacher began the session by handing out a new story which the students had not yet read. After previewing the book, both teachers had students take turns reading small sections of the text. After each section the teacher paused to clarify difficult concepts, explain vocabulary, and to ask questions. The teachers asked very few questions aimed at simply testing low level comprehension. Instead, the teachers asked questions to elicit students’ prior knowledge, text to text connections, and predictions. Sometimes these questions also led to more esthetic reactions about how students felt about the story. The teachers sometimes called on students who raised their hands and sometimes called on students who had not yet spoken. The teachers used the IRE format in which the teacher controlled talking turns, took every other turn, and uttered the majority of the words. However, in both class the small group format and patient but firm teacher control resulted in every student participating multiple times by the end of the 20 minutes session. The following is an example of a typical episode in the discussion of the first class:

Teacher: Good job. So elephants were used. What do we use now to haul things?

Student Two: Machines.
Teacher: Yeah, what kind of machines? If we want to get something really heavy from one place to another how do we do it now? What kind of machines do we use?

Student Two: Plows.

Teacher: Plows. Oh to plow fields. Yeah, that’s one thing elephants were used for. Have you ever been on the highway? Have you been on there and seen big big trucks with big heavy loads on them. We use trucks now. They go a lot faster, but we used to use elephants. Can you imagine that? What if you were riding down the road and saw an elephant pulling a big tree? What would you think?

Student Three: That elephant better get back in the zoo.

Teacher: That elephant better get back in the zoo. [laughs] That would be really uncommon. That would be kind of scary. Ok, do you want to read?

In this except, the students uttered only a few words compared to the teacher who spoke at length. The questions drew out student’s there prior knowledge so that it could be used to better understand the story.

The other teacher using the small group discussion elicited some prior knowledge but concentrated more on making inferences and eliciting feelings about the text. Here is a typical exchange that took place in this class (It should be noted that the story never directly states when Oscar took the batteries, so the first question is an inference rather than a simple recall question):

Teacher: I asked you earlier about what day of the week it was in the story. When do you think Oscar probably took the batteries?

Student One: Right after they used it.
Teacher: Ok so after they used it. Which night?

Student One: Sunday

Student Two: Saturday

Student Three: In the morning

Teacher: Ok it could have been Saturday morning or what?

Student: Friday

Teacher: Friday night because they said they used them the day before or actually the night before. So he either removed it after they used it or as someone said Saturday morning. OK now this one is more your opinion. Do you think that Oscar’s behavior is realistic for the age that he is. It says he is a five year old. So do you think he’s behavior is pretty realistic for what took place.

Student One: Because I have a five year old sister and she always gets into stuff.

Teacher: OK

This teacher had a very similar content and format to the first teacher. The only differences were that the second teacher did not require the same degree of turn control as the first teacher. As can be seen from the above example, students often had consecutive turns and the teacher sometimes responded with very short and neutral feedback, but the vast majority of the words were still uttered by the teacher and the teacher strongly controlled the topic of discussion.

The other two classrooms used whole class discussion. Despite the different format, the question types were very similar to the questions used in the small group classrooms. Students were asked a mixture of questions eliciting inference, prediction, connections with other texts
and prior experience, and personal feelings. The fourth teacher was unusual in that she had her questions written out on cards which she distributed to pairs of students. The students then asked the rest of the class to give answers. This disrupted the usual IRE format to some degree and gave more talking turns to the students, but the students asking the questions did not give much extemporaneous talk beyond reading the questions and the teacher still gave feedback to the students by interjecting frequently. Here is a brief example of one set of questions asked by a pair of students:

Student One: [Reading from card] The question is what has Ms. Williams and Little Willey’s teachers said about Willey?

Student Two: What have teachers said about you?

Student One: [Calls on student three to answer]

Student Three: Ms. Williams said, “That boy asks too many questions.”

Student One: What has teachers said about you? [Calls on student four]

Student Four: Well teachers in the past said I was a good student.

Teacher: Anything else? [pause] Oh come on, what do you think us teachers talk about in the teacher’s lounge?

In summary, the contrast between the CR and regular classrooms discussions largely held to the expected pattern. Teachers delivering their regular classroom discussion avoided some of the low level recall questions expected from some prior research in favor of higher quality questions directed at guiding students through good reading comprehension strategies and engaging them with the text, but the format was the traditional IRE with some variation between classes. In no case, were students in regular classroom discussion truly in charge of turn taking or topic control. Furthermore, no discussion was primarily about students’ opinions. While their
opinions were elicited at some points, the primary purpose of these questions was greater comprehension of the story, not the students’ development of argumentation. A good example of this is the second teacher’s question about whether the story was realistic. While he elicited answers, his intention was to demonstrate that the story was intended as moral fable rather than a “slice of life” type story. So this series of questions was leading the students to understanding the intentions of the author rather than their own ideas.

The observation of the discussions supports the expected difference in social accountability between the discussion treatments. In the regular classroom discussions, students were accountable for comprehending the story, but they were not accountable for having coherent opinions about the topics. Furthermore, the format of the regular discussions made students accountable for only very short isolated responses as compared to CR classrooms that featured long stretches of dialogic discourse.

**Reading time**

**Trimming and adjusting reading times.** Reading time data was collected from 127 students. One student was removed because he was not able to finish the reading, and another three students were removed because they had extensive problems following the directions for reading on the computer. The reading times for the 123 remaining students were first analyzed for unreasonably short times. One possible explanation for extremely short reading times was that some students accidently pressed the space bar twice which caused them to skip clauses. In these cases I chose to remove these reading times, because they did not reflect the student’s depth of processing and the extremely short values would have an unduly strong influence on the statistical conclusions. However, I did not want to exclude measurements for students who were skimming the text. Skimming is a reading procedure that does reflect the student’s depth of
processing and would be a clue to a student using a peripheral processing of the text. For this reason, I used a very conservative criteria based on previous self-paced reading research.

Previous data has shown that during self-paced reading, college students read at about 400 ms per word (Rayner & Clifton, 2009). Based on this data, I determined that any reading time for an entire clause that was below 400 ms precluded the student from having read even one word, and therefore they would have been unable to gain any information from that clause even if they chose to skim the clause. Using this criterion, I removed the reading times from 154 clauses (1.1% of all clauses).

Most students who had extremely short reading times had a very small number of them (M = 1.9; Mdn = 0). However, 16 students had more than 10% of their reading times in this range (11 or more clauses). Furthermore, these very short reading times tended to occur in clusters suggesting that students were attempting to skip portions of the story rather than individual clauses. It is unlikely that these students accidently skipped these clauses. A more likely explanation is that these students lost motivation to complete the task. These students were identified to allow for closer examination in later analysis.

Once the extremely short times were removed, extremely long reading times were trimmed. The remaining times were used to compute the mean per syllable reading time and standard deviation for each student. Per syllable reading times above five standard deviations from the mean for any particular student were replaced with a time equal to five times the standard deviation above the mean for that student for the syllables in that clause. This very conservative standard was used to take into account the many difficulties children experience during reading including unfamiliar words and concepts while adjusting for distractions such as other students speaking or students moving in their seats. This procedure resulted in trimming 80
reading times (0.66% of the total cells). Combing the reading times removed for being too short and times trimmed for being too long there were 234 altered times or 1.7% of the total number of cells.

**Analysis of reading times.** The reading times were analyzed using a mixed-effects model with crossed random effects for students and clauses. Many previous research papers analyzing research times have successfully used multiple regression using a two-step procedure (Stine-Morrow, Millinder, Pullara, & Herman, 2001). In the first step, the reading item level variables such as number of syllables and word frequency are used to predict each individual’s reading times using multiple regression. This produces regression estimates for each person. Then the regression estimates are compared to test individual level differences. This procedure addresses the fact that multiple reading times are not independent, because they are read by the same individual so that each reading time is influenced by the same variation of the reader such as her baseline reading fluency. The two-step regression procedure addresses this problem, but statisticians have argued in recent years that multiple regression cannot fully model multilevel variation because it is done in two steps. They have argued that mixed-effects models such as hierarchical linear modeling (HLM) give more reliable results with higher power because they analyze the variance from both sources in one step. As Richter (2006) explains, “The two-step [multiple regression] approach fails to separate the variance components pertaining to sentence and person level in an appropriate way because these variances are considered sequentially, not simultaneously as in hierarchical linear models with random coefficients” (p. 227). Examining the variance sequentially increases the number of error estimates. Multiple regression requires that the reading item level errors be calculated for each person, but in mixed-effects models they are calculated only once. This increase in error estimates lowers the reliability and power of the
analysis. In most cases, this difference is small, but Richter showed that for a sample set of reading time data, while the two-step multiple regression technique and HLM had almost identical reading item level results, multiple regression failed to find some of the significant effects at the person level found by HLM.

Given that the main research questions of this study are focused on student level variables, a mixed-effects model was chosen for this analysis. However, there are many types of mixed-effects models. Richter (2006) and many others have concentrated on hierarchical models to illustrate how to analyze reading time data. In these studies, different participants often get different item which are counterbalanced across subjects. In this way it is preferable to think of items being nested within students, because each item is only read by certain students. Therefore, the item level variance must be analyzed in terms of the participant level variables of only those people who are reading those items. In contrast, this study had each student read the same clauses in the same order. Since each person read each clause, it is more accurate to measure the variance of each clause across all the students. In this type of model, the participant and clause level are thought of as crossed instead of nested, and both levels are allowed to vary randomly (Baayen, Davidson, & Bates, 2008).

Due to the heavily skewed nature of reading time data, a transformation is necessary to make the data approximate a normal curve. A graphical analysis suggested that the reading times approximated a log linear distribution. The standard test for normality, the Kolmogorov-Smirnov test, was not useful in testing whether this transformation was appropriate, because it is almost always significant given a large sample size (n>100). However, the frequency and Q-Q plots shown in Figure 8 illustrates that the natural log of the values closely followed the expected
values except for at the extremes of the distribution. Therefore, the natural log of the reading time was entered as the dependent variable for this analysis.

One further note should be mentioned. It might be argued that students and clauses are further nested within classrooms. This more complicated analysis would more accurately model the research design in which students are grouped into different classrooms and the different discussion types (CR or regular) are assigned to whole classrooms rather than individual students. While this is true, I have chosen to ignore this third level of variance due to limited sample size. Unfortunately, research has shown that variance estimates are particularly sensitive to the sample size at the highest level (Mok, 1995). While this study has 112 clauses and 123 students, there are only 8 classrooms. This great disparity in the number of classrooms per student makes estimates at the third level unreliable. Therefore, one limitation of the study will be that discussion type will be modeled as a person level variable thereby ignoring the classroom differences which are undoubtedly present.

**Clause level variables.** Clauses were coded for variables at each of the three levels of discourse representation. However, because the reading times were recorded for clauses, it is difficult to distinguish between surface level variables and textbase variables. Summing or averaging across words makes it difficult to distinguish between aspects of individual words as compared to larger pieces of textbase meaning. Therefore, the variables associated with the surface level and the textbase level will be treated similarly in the analysis and no attempt will be made to separate their effects. The variables associated with the surface level included in this analysis are the number of syllables in each clause, the number of words in each clause, and the average frequency of the words in the clause. The average word frequency was calculated using the log of the number of Google hits for each word. Google hits was used instead of traditional
frequency tables, because Google hits are consistent with traditional methods while providing frequency estimates for names, slang, and contractions not usually available using traditional methods (Blair, Urland, & Ma, 2002). In addition, both the average frequency for all the words and the average frequency for content words were calculated. Content words included all the words except for prepositions, articles, coordinating conjunctions, and demonstrative pronouns. Based on previous research, it was expected that reading time would increase for clauses with more words and syllables (Rayner & McConkie, 1976), and that reading time would decrease as the average frequency of the words in the clause increased (Inhoff, & Rayner, 1986).

The number of propositions in each clause was added to represent the textbase. Based on previous research, it is hypothesized that reading time will increase with the number of propositions (Kintsch, & Keenan, 1973). Propositions were calculated using CPIDR version 3.2 with speech mode turned off (Brown, Snodgrass, Kemper, Herman, & Covington, 2008; Engelman, Agree, Meoni, & Klag, 2010). CPIDR is a software program that calculates numbers of propositions based on the number of words and their parts of speech. The software has shown high reliability with human raters (r=0.97 with two human raters), and it gave similarly high reliability with the examples published by Turner and Greene (1977).

The variables intended to reflect the variation in difficulty at the situation model level include serial order of the clauses, beginning of new arguments, argument wrap-up, and serial order of arguments (Britton, 1994; Stine-Morrow, Milinder, Pullara, & Herman, 2001). Serial order of the clauses was a sequential numbering of the clauses from one to 112. The beginning of a new argument was a dummy variable coded so that the clauses in which there was change in the argument being made (e.g. from an argument about how the floor could look flat when the Earth was round to an argument about how being too close to the Earth explained why it appears
flat) was coded with a one and all other clauses were coded with a zero. This often corresponded to a change in speaker due to the dialogue structure of the passage. Argument wrap-up coded the last clause in each argument with a one and every other clause with a zero. Serial order of arguments was a variable that coded the first clause in each argument as a one and then sequentially numbered the following clauses until the argument being discussed changed.

There are two additional clause level variables that do not fit neatly in the three level structure that has been discussed. These include a dummy code indicating the topic of the clause and dummy code indicating clauses after pictures. The after picture variable is a dummy variable that marked each clause that appeared directly after a picture with a one and all other clauses with a zero. This variable was intended to identify and control for any unintended variation due to the insertion of pictures into the story. The topic variable categorized the variables into those clauses that argued for and against a round Earth. Clauses arguing for a round Earth were given a one, and clauses arguing against a round Earth were given a zero. Eleven of the clauses consisted of exposition that was neutral to the question of the shape of the Earth (e.g. “Jill and Maria are talking after school”.) and were clustered at the beginning and the end of the story. These clauses were not coded and were entered as missing data for this variable. A separate analysis of these clauses was not attempted, because they were not considered to have any impact on the research questions of this study.

Student level variables. Student level variables include the four pre test measures (reading fluency sentence verification test, cognitive abilities figure analysis subtest (CogAt), the need for cognition survey (NFC), and the Stanford Abilities Test 10th edition (SAT-10)), age, gender, discussion type (CR or regular), whether the student received instruction that included an
Model selection. The best model was selected using a blockwise selection process (Pedhazur, 1997). In this procedure the variables are grouped into categories. The variables are entered block by block. Since many of the variables at the same level were correlated (see Table 4) not all the variables could be included. Based on predictive power, multicollinearity, and theoretical significance a subset of the variables in each block are chosen, and the remaining variables are removed. This process is repeated for each block. This process ensures that the final model will consist of representatives from the theoretically important categories. As with any selection process, the selection of variables is based on predictive rather than explanatory power; therefore, the treatment variables were kept in the model regardless of significance as they are necessary to test the central research questions.

The first block of variables to be entered in this analysis were the demographic variables of age and gender. Neither of these variables were significant predictors. The four pre-test scores (CogAt, fluency, SAT-10, and NFC) were then added. Of these scores only fluency was retained. Both the SAT-10 and fluency would be significant predictors if entered individually, but due to significant correlation between these two measures only one could be retained. Fluency was chosen because it is a stronger predictor and has greater theoretical connection to reading speed.

After the student level variable of fluency was established the clause level variables were entered starting with the variables associated with the surface and textbase representations of the text (number of syllables, number of words, number of propositions, average frequency of the all the words, and the average frequency of the content words). The first three variables are all measures of length and are all highly correlated; therefore only one could be entered to avoid
 multicollinearity. Number of propositions was chosen because it had high predictive value and had theoretical significance. Neither of the frequency measures were significant predictors. This is not surprising as the variance due to word frequency is difficult to measure over multiple words in a clause.

The variables associated with the situation model (serial order of the clauses, beginning of new arguments, argument wrap-up, and serial order of arguments) were entered next. The three argument based variables are highly correlated, so only one could be entered. New argument was chosen, because it had the highest predictive power given the variables already in the model. Serial order was also retained, because it has a separate and significant predictive effect.

Next the two additional variables were entered (after picture and clause topic). The after picture variable was not significant given the variables already present. This is likely due to the fact that the clause after a picture often began a new argument; therefore, the new argument variable already explained most of the variance caused by the pictures. There was also no significant difference between clauses arguing for or against a round Earth.

Finally, I entered the treatment variables of type of discussion and announcement. I also added various interaction effects.

In summary the final model consisted of the clause level variables of number of propositions, new argument, and serial position, and the student level variables of fluency, discussion treatment, and announcement treatment. In addition, the model had the theoretically important crossed level interactions of each treatment variable with the number of propositions, new argument, and serial position.
**Results for reading times.** The main analysis of reading times was conducted using only those students who skipped less than 10% of the clauses. The results shown in Table 5 shows that both the student variable of fluency and the clause variables of number of propositions, new argument and serial order all had significant main effects. As predicted by previous research, reading times increased for clauses with more propositions and for clauses that marked the beginning of a new argument. Reading times declined as students moved through the text. Furthermore, reading times decreased for students with greater fluency.

For the treatment conditions, only the discussion treatment produced significant changes in reading speed. The announcement treatment had no significant main effect or any interactions. The positive beta weight of the discussion treatment indicates that reading times were longer for students who participated in CR discussions than for those students who had regular language arts instruction. Furthermore, the positive significant interaction between propositions and discussion shows that students who participated in CR discussions slowed down particularly for clauses with more propositions. However, the CR and regular instruction students experienced similar reading time increase when reading clauses with new arguments and similar reading time decreases as they moved through the text.

**Analysis of reading times for students who skipped more than 10% of clauses.** The 16 students who skipped more than 10% of their reading times have too much data missing to be added to the general reading time analysis. However, they also provide a source of information about the reading behavior of the students in the study, and they are numerous enough to allow for some statistical analysis.

To investigate why some students might choose to incompletely read the text I created a dummy variable indicating those who skipped more than 10% of the clauses. I then used logistic
regression to predict this variable using age, gender, the four pretest measures, and the treatment conditions, and a series of dummy codes categorizing their model of the Earth prior to the reading.

The results listed in Table 6 indicate that students who skipped a large number of clauses tended to be slightly younger and mostly male (11 male, 5 female). This indicates that part of the issue might be a certain lack of maturity (no offense to boys). After accounting for these variables, there are still effects for spatial reasoning, reading comprehension (as measured by the SAT-10) and need for cognition. The result for fluency should be treated with caution, because it is highly correlated with SAT-10. Plotting the need for cognition score by the SAT-10 score for both groups reveals that the regression line for students who did not completely read the story have a larger slope due to a number of students near the lower left-hand quadrant and a few students in the upper right hand quadrant (Figure 9.). Those students in the lower left-hand quadrant have lower than average need for cognition suggesting that they generally do not like challenges, and lower than average reading comprehension indicating that reading the text might have been challenging. These students might have skipped clauses out of frustration because the reading was difficult, and they do not like difficult tasks. Students in the upper right-hand quadrant have higher than average need for cognition suggesting they like challenges and higher than average reading comprehension indicating that reading the text was easy for them. They might have skipped clauses because they were bored with a task too easy for them. This hypothesis is bolstered by the fact that these students were more likely to have had pre-test models of the Earth that passed the strict test of consistency to the sphere model.

Finally, the students that skipped clauses tended to have had CR training and had no announcement of a future discussion. This might indicate that having an announced future
discussion, especially if it might mean accountability directly to the teacher for text comprehension was one incentive for some students to not skip clauses even if they found the reading too simplistic.

**Analysis of the Sentence Verification Test of Comprehension (SVT)**

The data from the SVT are the cumulative score of fifteen true and false questions. This results in a score that is not normally distributed. The distribution has a negative skew ($M = 11.078$, $SD=2.73$, skew=$-0.56$) resulting in a significant Kolmogorov-Smirnov test for non-normalcy ($p=.05$). The negative skew is a result of the difficulty of the test, but it is also caused by the true and false format which naturally results in a 50% accuracy rate due to guessing. A square root arcsine transformation was performed to improve the approximation of the data to the normal distribution. The transformed data met the Kolmogorov-Smirnov test ($p=0.25$). This is also represented in the histograms and Q-Q plots (Figure 10).

As with reading times, model selection was conducted using a blockwise selection technique. First the demographic variables of age and gender were entered. Neither of these variables were significant predictors of SVT scores. The next block consisted of the four pretest scores and a dummy variable indicating if the student’s pre-test model of the Earth was consistent with a spherical Earth. The only variable in this group that is a significant predictor is reading comprehension as measured by the SAT-10. This is expected since reading comprehension ability in general should predict the reading comprehension of the specific text. Finally the treatment variables and their interactions with the SAT-10 were entered.

The results shown in Table 7 show that the discussion treatment did not predict SVT scores, but the announcement treatment was a significant positive predictor. This means that students who received an announcement of a future discussion scored higher on the SVT than
students who received no announcement. Furthermore, there was a significant negative interaction between SAT-10 scores and the announcement condition. This interaction was explored using Figure 11. Figure 11 shows that students with low reading ability did better on the SVT when they were in the announced condition, but this effect did not exist for students with high reading comprehension.

**Analysis of Self-Assessment of Processing**

The students answered twelve Likert scale questions designed to measure the students’ own assessment of how deeply they processed the text. In general, the students gave high ratings. The mean for all the questions except for the negatively worded question 10 had ratings above the mid-point of three. Students overall found the story interesting (M=3.75, SD=1.28) and easy to understand (M=3.95, SD=1.16) However, all the questions showed a full range of answers and there was significant variation among the students (Table 8).

While all the questions were intended to address the single underlying construct of depth of processing, the questions were drawn from a variety of sources, and they addressed a variety of different aspects of deep processing including the use of metacognitive reading strategies (i.e. When I was reading this story, I sometimes stopped reading to think a little while before I kept reading.), interest (i.e. The story was interesting to read.), and effort (i.e. I put a lot of effort into reading and understanding this story.). The variety in these questions made it prudent to begin a data reduction analysis to determine the extent to which the different questions constituted a single measure.

The data reduction analysis was conducted using the principal component method. Two factors emerged with eigenvalues over one (Table 9, and Figure 12. For scree plot). The factors together explain 40% of the variance. To ease interpretation of these factors the factor loadings
were rotated using the Promax method. Factor loadings for each factor are listed in Table 10. Promax is an oblique rotation, meaning that the rotation does not require that the resulting factors are orthogonal. This technique was chosen, because the questions are assumed to measure correlated factors instead of distinct aspects of deep processing. Furthermore, the Promax method produced factors that were more easily interpreted than orthogonal methods of rotation such as Varimax.

To interpret the results of the Promax rotation, I used the rule of thumb that factor scores greater than 0.30 are significant. Since Promax rotation is an oblique method, I also allowed for split loadings (questions loaded onto two factors). The first factor which consists of the following items:

1) I think I understood this story pretty well
2) The story was interesting to read.
6) Reading this story made me think a lot about the shape of the Earth.
7) When I was reading this story, I asked myself if I really believed what I was reading.
8) Knowing about the shape of the Earth is important to me.
9) When I was reading this story, I tried to picture in my mind the things I was reading about.
12) When I read this story, I thought about what I might say during my ____ discussion.

I interpreted this factor as primarily a measure of engagement in the story and the story topic. This factor includes questions having to do with interest in the story (question 2) and interest in the topic (question 6 and 8). It also includes questions asking students if they used strategies of engagement during the reading such as thinking about the topic while they read (questions 6, 7, 9, and 12). Students who were interested and highly motivated (or wanted to appear motivated) to read the story would give high ratings to these items.

Factor 2 had items with both positive and negative loadings. Therefore, factor consists of the contrast between the following items:

3) I put a lot of effort into reading and understanding this story.
5) When I was reading this story, I sometimes stopped reading to think a little while before I kept reading.
6) Reading this story made me think a lot about the shape of the Earth.
7) When I was reading this story, I asked myself if I really believed what I was reading.
10) When I read about Jill and Maria and they did not agree with each other, it made me confused.

Versus

1) I think I understood this story pretty well
4) This story reminded me of other things that I’ve already read and heard about.

I interpret this factor as primarily a measure of the perceived difficulty of the reading. This factor includes global judgments of difficulty (questions 1, 3, and 10). It also asked students if they used strategies needed to understand difficult texts (questions 4, 5, 6, and 7). While these strategies are used by good readers frequently, they are more likely to be salient to students who had just read a difficult text. Furthermore, students might be more likely to admit to using these strategies if they want to express that they found the text difficult.

Question 11 does not load onto either of the factors. Question 11 has by far the lowest communality estimate (0.0368) of any of the questions (the next lowest is question 5 having a communality of 0.190). This indicates that question 11 shares little variance with the other questions and is measuring something unique. Question 11 states “When I read about Jill and Maria and they did not agree with each other, I chose one of them to believe.” This question is not about interest or effort at reading. Rather it measures the student’s effort in thinking about their own opinions on the topic in relation to the text.

Four separate regression equations were calculated in this analysis. One regression equations was calculated for the total self-assessment score, and one regression equation was calculated for each factor. Finally a regression was calculated for question 11 alone. As with the analysis of the SVT, each of these regression models was selected using the blockwise selection
process. The demographic variables of age and gender were entered followed by the four pretest scores, a series of dummy variables representing the student’s pre-test model of the Earth, the treatments, and finally interactions.

The final model for predicting the total self-assessment score consisted of the need for cognition score, a dummy variable describing if a student had a spherical pre-test model or not, and both treatment variables. There were no significant interactions. Table 11 shows that need for cognition survey was a significant predictor of student’s self-assessment. This is expected, since students who like to think are likely to be more engaged and try harder to read a text about a scientific controversy unless they find the controversy too easy. Those students who found the text too easy are reflected in the large negative significant estimate for the spherical variable. This shows that students who already had a spherical model rated themselves as processing the text less deeply than students who did not already have a spherical model and who must have found the text more controversial and challenging. In addition these results show that students in the CR condition felt that they did not process the text as deeply as students who had regular discussions. This will be explored more through the following analyses.

As described above, factor one centered on the student’s engagement with the text. The final model for predicting factor one was identical to the model for the total score. Table 12 shows that factor one had similar results to the total score. As before, need for cognition had a significant positive effect, but having a spherical pre-test model had a large negative effect. Both the discussion treatment and announcement treatment had the same direction as the total score, but each had somewhat weaker effects. For factor one, the discussion treatment was only marginally significant effect.
Factor two is centered on the student’s judgment of the text difficulty. The final model for predicting factor two was identical to the model for the total score, except that instead of need for cognition, the SAT-10 reading comprehension test was a significant predictor. This is consistent with the interpretation of factor two as a measure of judgment of text difficulty. Table 13 shows that SAT-10 had the predicted direction so that students with higher reading comprehension found the text less difficult. Also predictably, students who already had a spherical model of the Earth found the text less difficult. The treatment variables had the opposite trends found in factor one. In this case, the discussion treatment is much weaker than the announcement treatment which is marginally significant.

The model predicting scores on question 11 was identical to the model for the total score. Table 14 shows only need for cognition was a significant predictor of question 11.

In summary, the results of the self-assessment show several patterns of results. Firstly, the students who rated themselves high on need for cognition six weeks earlier rated themselves as processing the story more deeply. This suggests that the need for cognition survey was an accurate predictor of how students will choose to engage in an intellectual task like reading about a science controversy. The predictive power of the need for cognition survey also shows that this personality trait has at least some generalizability across academic situations. Secondly, students who had a spherical model of the earth at the time of the pre-test five weeks before the reading gave lower self-assessment ratings overall for both factors of engagement and difficulty. One possible explanation for this is that students with a spherical model had more prior knowledge and therefore found the text easier and less engaging. The announcement treatment did not have an overall effect on the self-assessment. In contrast, the discussion treatment was a significant predictor such that those students who experienced collaborative reasoning reported less depth of
processing. The analysis of the two factors suggests that this effect resulted mostly from a

decline in engagement rather than a feeling that the text was easier. The results using question 11

as the dependent variable further show that this decline in self-reported depth of processing was

not true for students’ assessment of how much they thought about their own opinions while

reading. This result is bolstered by the means of each question separated by discussion

treatments (Table 15). CR students had lower means on all of the items except for two: question

9 and 11. Both these questions refer to students’ inner thinking and are associated with both the
text and students’ own thoughts. Another way of examining the relative importance of this

question controlling for the general decline in engagement for CR students is to rank order the

items from the highest ranked to lowest ranked for students in each discussion treatment. For

students who had regular instruction, question 9 was moderately ranked and question 11 was one

of the lower ranked items (ranked 4th and 9th respectively), but for CR students both questions

were very highly ranked (ranked 1st and 3rd respectively).

Analyses of the shape of the Earth interview

Out of the original 130 participants, there are 102 completed and coded pre-test

interviews and 96 post-test interviews. In addition to absences, some students’ interviewers were

unusable due to failures in recording technology or inadvertent interview errors such as skipping

essential questions. The results for the pre and post test interviews are summarized in Table 16.

To compare these results to the original data observed by Vosniadou and Brewer (1992),

Table 17 collapses the loose and strict models and presents the percentages of students in each

model. In Table 18 is the comparative data from Vosniadou and Brewer (1992) collected for

third and fifth grade students. The largest difference between the results of the present study and
the results of Vosniadou and Brewer is the greatly inflated number of students with mixed
models in the present study. As explained earlier, this might be due to different interview procedures.

**Change in models.** Since a student required both a complete pre and post test interview, only 83 students were available for this analysis. Table 19 summarizes the results collapsed across both treatments. In the table, the student’s model at the beginning of the experiment is listed on the left. The student’s model after the treatments and reading the target text is listed on the top. The models are ordered according to the developmental order presented by Brewer (2008). In addition, each model was subdivided into strict and loose consistency with the model scoring rubric. Models that had inconsistent answers and were categorized as mixed were given the lowest position in the order. In Table 19, progress is defined as any movement to a more sophisticated model or a move from a loose model fit to a strict fit for that model. Students who had no change can be seen along the diagonal, students who progress are below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics). Using this scheme, students who began with a strict model cannot possibly progress. Students who began with a mixed model cannot regress.

Overall 44 students (53.0%) progressed, 25 students (30.1%) had no change; and 14 students (16.9%) regressed. Of the 19 students who began with spherical models that only loosely fit the scoring rubric, 16 of these students ended with a spherical model that strictly fit the scoring rubric. Three students started with hollow models of the Earth and ended with spherical models of the Earth. One student with a dual Earth model adopted a hollow Earth model. Of the 31 students that began with a mixed model, 7 progressed to a hollow Earth model and 16 progressed to a spherical model. Of the 14 students that regressed, 8 of them were classified as having mixed models.
The following tables will present the conceptual change results separated into the different treatment groups. These tables will display the information to give readers the exact results. However, any differences between groups will be further analyzed using a logistic regression analysis in a later section.

Table 20 and Table 21 present the conceptual change results for each discussion treatment. Due to unequal absences, class sizes, and other causes, there were more complete sets of interviews for students from the CR condition than the regular instruction condition (n= 53, n= 30 respectively). In addition, a greater percentage of CR students began with spherical models of the Earth that strictly fit the scoring rubric (12 CR students, 1 Regular Instruction student). Despite these differences, the percentages of progressions and regressions between the conditions were similar (For CR progressions = 27 or 50.1%, regressions = 8 or 15.1%; for regular progressions = 17 or 56.7%, regressions = 6 or 20%).

Dividing the results according to the announcement treatment produced almost equal groups (n= 42 for Announced, n=41 for unannounced) (See Table 22 and Table 23). However, there were more students who began with strict spherical models in the unannounced treatment groups (5 students in the announced group, 12 in the unannounced group). The percentages of progressions and regressions were similar in both groups (For the announced group progression = 21 or 51.2%, regression = 9 or 22.0%; for the unannounced group progressions = 23 or 54.8%; regressions = 5 or 11.9%).

Dividing the results in terms of the immediate and delayed post tests produced similar sample sizes (immediate = 36; delayed = 41 (See Table 24 and Table 25). In this case, more students in the delayed post test condition began with strict spherical models (immediate = 4; delayed = 13). There were also a somewhat higher percentage of progressions in the immediate
condition (Delayed progression = 20 or 48.8%, regression = 7 or 17.1%; immediate progressions = 22 or 61.1%, regressions = 4 or 11.1%).

**Analysis of model change using logistic regression.** Logistic regression is a statistical model using general linear modeling to predict the probability of an event occurring (Agresti, 1996). Logistic regression requires a binary dependent variable; it assumes that for each participant the event either did or did not occur. Logistic regression fits the probabilities of the dependent variable to the logistic curve. In the case of this analysis, the model will be predicting the occurrence of students adopting a better model during the post-test interview than the one they had at the time of the pre-test interview. Since it is impossible for students who already have a strictly spherical pre-test model to improve on the post-test interview, they were not included in this analysis. This left a total sample size of 70.

**Model Selection**

As with previous variables, this model was selected using the blockwise selection technique with the order being age and gender, than the four pre-test measures, a group of dummy variables representing the students’ pretest model, a variable representing the timing the post-test interview (immediate or delayed), the treatments, and finally interactions. This produced a model consisting of the student’s Cognitive Abilities Test score (CogAt), a dummy variable representing if a student’s pre-test model was mixed or coherent, the discussion treatment, the announcement treatment, and a discussion treatment by mixed pre-test interaction. The time of the post-test interview had no main or interaction effects and was therefore removed from the analysis.

**Regression results.** The results in Table 26 show that student’s scores on the CogAt are highly predictive of their probability of progressing to a more sophisticated model. Students who
progressed had CogAt scores almost one standard deviation above students who did not (M = 9.60, SD=3.37; M=6.38, SD=3.72 respectively). This suggests that the ability to mentally manipulate objects helped students to benefit from the text. This is reasonable, because the text relies heavily on the description of the Earth from different perspectives.

The results also show a significant result for the variable representing whether students had a mixed or coherent pre-test model. This suggests that students who began with a mixed model were more likely to progress than students who began with a coherent model (21/39 or 54% of coherent students progressed; 23/31 or 76% of mixed students progressed). The fact that such a high percentage of students with a mixed model progressed suggests that the inconsistent nature of the answers of students with mixed models is more indicative of students in transition between models or students who have poorly articulated models rather than students who have firm but incoherent beliefs.

After accounting for spatial reasoning ability and pre-test models, there was no main effect for either treatment. However, there was a significant interaction between students having a mixed pre-test model and the discussion treatment (Table 26). Figure 13 illustrates the effect of this interaction. For students who had coherent pre-test models, the CR treatment lowered their rate of progression compared to students receiving regular instruction (12/24 or 50% and 9/12 or 62% respectively). For students with a mixed pre-test model, CR treatment increased their rate of progress (15/17 or 89% and 8/14 or 57% respectively). This interaction suggests that the experience of argumentative discussion caused students who already had a coherent model to become somewhat more entrenched in their beliefs, while students who did not have a coherent model were strongly encouraged to use the text and their own ideas to coordinate their ideas into a coherent model.
Examples of conceptual progress. This section will present two examples of students’ pre- and post-test interviews to illustrate the range of conceptual growth encompassed by the model progress measure used in the logistic regression. Some students who were categorized as progressing in their models had changes in their interviews resulting from changes in belief about specific issues as well as improvements in their articulation of their models. Other students seemed to have more profound changes. A more carefully analysis of the interviews is planned for the future to measure the percentages of these different types of changes across students and treatments. However, given the relatively short duration of the intervention it is likely that that majority of changes constituted a more incremental conceptual growth rather than a radical conceptual change.

Incremental change example. This student showed enough progress to move from a mixed model to a loose spherical model. This student started having a mixed model that had elements of a spherical model but also some inconsistent answers. At the post-test he showed progress, but he still had some difficulties with the issues. This amounts to an incremental change in which this student has not radically changed to a completely new model, but he has developed a more coherent and consistent model.

At the pre-test, this student answered the first question in a way that seemed most appropriate for a disc shaped model of the Earth:

Student: It’s not a plate circle, but it’s kind of a circle. Like it’s a circle, but there’s bumps and stuff on it.

Interviewer: Okay. What are the bumps on the circle?

Student: How high some of the grasses and stuff go…

Interviewer: Like mountains?
Student: Yeah.

This same student then answered question eight in a way that suggests some type of hollow earth model, because it refers to an Earth which is round on the outside and flat on the inside:

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?

Student: [Nods Yes]

Interviewer: How come the earth is flat here, but before you said it was round?

Student: Because if you go down lower, it’s round and the gravity pulls them down onto the earth, so…

Interviewer: If you go down and it’s round, does that mean, if I go down from the house, or I were to keep...

Student You can see it round on here, but if you were more down low on Earth then it would seem flat to you, because the gravity pulls you down.

Despite these spherically inconsistent answers, this student also gave some answers which are very diagnostic of a spherical Earth including saying that China would be on the opposite side of his drawing from Illinois, and that people can live on the bottom because gravity pulls them down.

On the post test, the student’s answers become much more consistent to a spherical model of the Earth although he still demonstrated some confusion. This incremental change appears in the first question, which the student now answers by simply stating that the shape of the Earth is round. Furthermore, on question eight he expounds on this in his explanation for why the ground in a picture looks flat:

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?
Student: Yes.

Interviewer: This house looks like it’s on flat land, if I look outside it looks like these houses are sitting on the flat land. How does it look like it’s flat land when you said the Earth is round?

Student: Because those are only little parts of the Earth and the clouds and stuff go around can make it look round. It’s not perfect circle. Some parts are round but that’s only kind of like the small pictures of it. The real shape is round.

Now the student has progressed. He still has not worked out the perspective issue well, but he has learned enough from reading the story to move to an acceptable deviation from the expected answer. He now thinks that some parts are round and some parts are flat. He has come to a new understanding of how the earth can be round and still appear flat that while not completely scientifically correct still shows great progress in coming to terms with this fundamental assumption.

**Radical conceptual change example.** The next example demonstrates that at least some students did have a radical shift in models between the pre- and post-test interviews. This student switched from a dual Earth model to a hollow Earth model.

The first clue to this radical shift are the differences between her pre- and post-test drawings (Figure 14). Her first pre-test interview had showed the moon, stars and clouds inside the sphere and people lined up in a line. Her post-test interview showed the moon, stars and clouds outside the earth and people distributed over the circle.

In addition, she demonstrated a change over all of her answers suggesting a general change in model rather than an incremental change in certain details. For example, her pre-test
answer to question eight is confused but indicative of a dual-Earth model because it distinguishes between space and land, and it acknowledges that at least one earth is flat.

Interviewer: Here is a picture of a house. This house is on the earth, isn’t it?

Student: Mhm.

Interviewer: How come the earth is flat here, but before you said it is circle?

Student: Because this is out of space, and space not everything get’s flat and in space is different from land because land is flat. Some lands are kind of hills but the earth is flat.

In her post test interview, she does not distinguish between two different objects one in space and on Earth. Now she describes two places on the same object, one of which is flat and one of which is round.

Interviewer: Here is a picture of a house. This house is on the earth, right?

Student: Mhm.

Interviewer: How come the earth is flat here, but before you said it was round?

Student: Because deep down in one end is flat and on earth there is lot of gravity and stuff in it and it’s different from when you are on earth. Cus it’s flat on the land, flat and you get up at the end of the earth it’s not flat cus it’s round.

A similar change is seen in this student’s belief about the end of the Earth. In her pre-test interview she expressed a belief that there was an end of the Earth and people call fall off into space if they went there.

Interviewer: Could some people fall off of the end, edge of the earth?

Student: Mmm – yes.

Interviewer: Where would they fall?
Student: They could fall, if they fall off the earth they could fall in outer space may be? And they wouldn’t be on earth, they will probably fall in outer space.

After reading the story, the student demonstrates a fairly clear hollow Earth model despite the fact that this model was never explicitly described in the story.

Student: You couldn’t walk to the end of the earth and then fall because you are not all the way in the space. You are on earth, you are not on top of the earth but you are like part of the earth so that you can’t fall off the earth because you are not like… in the space where you can fall.

The final example of this student’s model change is given in the last question. In the pre-test interview the student explains why people cannot fall off the bottom in terms of two places one in outer space and one that is the actual earth.

Interviewer: Now you said nobody lives here, but do people live, let’s say: here or here?

Student: Yes, cause like people live in south and north and stuff.

Interviewer: How can they walk there for a few days without falling off?!

Student: Because they are not, like in outer space they are in a country, they are not like they are living on the actual earth, they are living on earth but they are living on homes

On the post-test interview the student does not contrast two different locations, but two parts of the same planet: the inside and the outside.

Student: They don’t fall off because they are not else-. They are kind of out off, not outside the earth, they are in the earth and when you fall you are walking in different places but you are not off the earth yet.
These examples give evidence that the students categorized as progressing from the pre- to the post-test represent a range of degree of changes which include at least some radical conceptual shifts as well as more subtle shifts in aspects of their models and in their articulation of their beliefs.
Chapter 5

Discussion

This experiment set out to test dual-process theories of conceptual change as its main focus and the effect of classroom discussion on depth of processing as a subsidiary goal. Dual-process theories of conceptual change contend that students will choose to process a lesson at various depths. Those students who process the lesson more deeply will be more likely to change their conceptions and that change is more likely to be complete and persistent over time (Dole & Sinatra, 1998). To test this theory an experimental manipulation is needed that can change the degree of processing of students. Classroom discussion was offered as a good means of doing this based on previous research (Chaiken, 1980), and this study was designed to investigate this effect in the classroom context.

To improve on previous attempts at testing dual-processing theories of conceptual change (Acee and Weinstein, 2010; Alexander, Fives, Buehl, & Mulhern, 2002; Hynd, 2003; Nussbaum & Sinatra, 2004), this experiment varied the motivation to engage with the text while holding the instructional materials constant. In addition this study set out to assess depth of processing through multiple methods including online and offline measures to directly test the predictions of the dual-processing theories as well as to probe the effects of discussion on processing.

The study was based on the hypothesis that the announcement of a future discussion and experience with CR discussions would have an individual and additive effect on depth of processing while reading the story which would be reflected in longer reading times, better comprehension, and a stronger feeling of effort and engagement. This deeper processing was further expected to lead to more frequent and more persistent conceptual change as reflected in a change in pre- and post-test interview answers, especially after a delay.
The actual results presented a more complicated picture than predicted. Reading times increased for the students having the CR discussion treatment, especially for information rich sentences, but the announcement treatment did not increase reading times. In contrast, the announcement improved students’ comprehension of the target text, especially for poor readers, but CR had no effect on comprehension. Most contradictory, students experiencing the CR discussions had lower self-assessments of depth of processing especially in the area of engagement, and the announcement had no significant effect on their self-assessments.

The effect of the treatments on conceptual growth also presented a complex picture. While the announcement treatment had no significant effect on improvements in students’ models, the CR treatment showed a complex pattern of effects. Students who began with a coherent model of the Earth and experienced the CR discussions showed somewhat less change than comparable students experiencing regular instruction. Students who began with a mixed model of the Earth and experienced the CR discussions showed a fairly strong increase in conceptual change compared to the other students. The persistence of this change could not be evaluated, because there was too little difference over the one week delay.

I believe that overall these results support dual-process theories while suggesting some important limitations and possible refinements. The theory’s central claim that depth of processing is a key determinant of how students will be persuaded by a lesson was supported by the finding that CR discussions successfully increased depth of processing during reading and also increased conceptual growth. However, there was not support for the contention that depth of processing is a single continuum with a single object of processing: comprehending the message. In the educational context of the CRKM, Dole and Sinatra (1998) describe the lowest level of processing as consisting of “simple strategies, such as maintenance rehearsal or
mnemonics” (p 121). They describe the highest level of processing as consisting of “elaborative strategy use, and significant metacognitive reflection” (p.121). Included in their definition of metacognitive reflection are both basic comprehension strategies like making inferences and personal connections as well as being “reflective about what they were thinking and why,” and thinking “deeply about the arguments and counterarguments related to the message” (p. 121). The degree of elaboration in this theory reflects complex qualitative shifts, but it is a unitary continuum and it always describes the elaboration of the message. However, the results of this experiment are not consistent with students moving along one continuum directed at the message, because students in the CR condition read more slowly and had greater conceptual growth under certain conditions, but did not have better reading comprehension scores and did not rate themselves as processing more. The theory does not clearly explain how these measures would diverge if they are all hallmarks of students who are reading with elaborate reading strategies and deep consideration of the text. The CRKM would predict that a treatment that increased reading time and lead to greater conceptual growth would also improve comprehension and would involve increased use of the very types of strategies asked about in the self-assessment survey.

The data from this experiment would be better explained by a theory which incorporated two objects of processing: the text and the student’s own beliefs. This view varies subtly from that of the CRKM because it allows students processing of their ideas and of the text to vary independently. Using this view, the results of this experiment are much easier to explain. CR discussions hold children accountable for having articulated opinions backed with reasons and evidence whereas regular classroom discussions hold children accountable for having better comprehension of the text. In the context of CR discussions, students will be more successful if
they devote processing capacity to their own beliefs in light of the text, so that they can use the text to bolster and coordinate their thoughts in preparation for expressing a cogent opinion and backing it with convincing evidence. This process of thinking about one’s own opinion, the text, and coordinating the two, would demand cognitive resources and would lead to the increased reading times seen in the data. This hypothesis would also explain the longer times observed for clauses with more propositions, as these clauses have more information with which to compare with one’s own ideas, making them particularly cognitively demanding. However, if students attempt to process the text in this dual-focus manner, their increased processing time would likely not improve comprehension as measured by the SVT, because the increased complexity of the task might interfere with basic reading skills such as monitoring for surface comprehension, and it would focus the student on the situation model rather than the detailed comprehension of the syntax and gist which are tested by the SVT.

In addition, the dual-focus-of-processing hypothesis would explain the increased conceptual change experienced by students who had CR training and began with mixed models despite not having improved reading comprehension. This is explained by the theory of van den Broek and Kendeou (2008) that conceptual change during reading is produced by co-activation of both the situation model presented in the text and the misconception of the student. To demonstrate this effect van den Broek and Kendeou had participants read regular science texts which presented only the scientific view or refutational texts which presented the scientific view and argued against a common misconception. Students with misconceptions spent more time comparing the theories as judged by a think-aloud protocol and had longer reading times judged by self-paced reading, but only if they were reading a refutational text. Van den Broek and Kendeou theorized that the refutational text facilitated conceptual change because it encouraged
participants to co-activate the scientific theory and the naïve theory in working memory so that they could be compared and coordinated. They used a computer simulation to show that the co-activation theory could account for the data. In this dissertation, all the participants read the same argumentative style text, but the CR students had greater motivation for co-activation. This would explain why students who were uncertain and had conflicting models were able to compare their models to the text and repair and improve those models if they had the spatial reasoning ability necessary. Students who already had coherent models would not be expected to gain the same benefit because they already had a cogent opinion to present during the discussion.

Finally, the hypothesis of dual-focused processing and co-activation of theories is consistent with the results of the self-assessment survey which show that CR students felt that they made less use of many reading strategies that target basic comprehension. These results are consistent with a cognitive resource allocation strategy that directed some cognitive resources away from reading the text. The hypothesis further suggests that these resources are instead used by students to co-activate the text and their own beliefs. Unfortunately, this contention was not tested by the survey, because the survey was not designed to measure processing on one’s own beliefs.

In retrospect, confusion about what is being processed — just the text or also the reader’s own ideas — was present in the original studies of dual-processing theories, including the study by Chaiken (1980) on the effect of discussion on depth of processing and opinion change. In these experiments the most widely used method of measuring depth of processing was to show participants a message and then ask them to list all the ideas they had about the message (given a time limit). These ideas were then coded for being relevant to the message and for being positive or negative about the opinion expressed in the message. However, the message-relevant thoughts
did not actually have to be about the message content itself as long as they were relevant to the message. For example, Chaiken lists as a message relevant thought the following statement, “The economic advantages of the trimester agree with me.” Since this statement is a comparison between the text and the participant’s own opinion, it is not clear how much the participant is processing the message and how much the participant is thinking about his/her own opinions. One can imagine that just as there is a standard of coherence while reading that determines how much cognitive energy is needed to reach the minimum level of comprehension of the text (van den Broek, 2001), there can also be a standard of coherence of one’s thoughts which determines how much mental energy will be expended to form one’s opinions into the minimum level of coherence.

**Role of Discussion in Motivating Science Learning**

The study was intended to evaluate the hypothesis that discussion is an effective method of motivating students to process a science text. In particular, it was hypothesized that experiencing CR discussions would have an especially large effect because CR discussions have a greater degree of social accountability than traditional formats.

The results did not show the expected additive effects. Instead the treatments had different effects. The announcement of a future discussion improved reading comprehension, but not performance on any of the other measures. Having experienced CR discussions increased reading time, increased conceptual change for students with mixed models, but decreased self-reported engagement. The lack of an additive effect and the strong main effects for CR even without an announcement suggests that students in the CR discussion anticipated a future discussion without being given the announcement. In hindsight, this is logical, because the CR discussions and the interviews were all conducted at the same times each week with the
researcher present. Furthermore, the story had the same argumentative format as stories students had read in preparation for CR discussions. These factors likely served as cues that a discussion would follow the story. This apparent anticipation by the students meant that most of the effect of the announcement treatment was likely due to the difference between the announced and unannounced regular discussion students. As such, the significant effect of announcement on comprehension matches the expected discourse. Announcing a future regular classroom discussion would motivate low readers to be more careful as they would be held accountable for comprehension.

The results that CR motivated students to exert more cognitive effort during the reading regardless of announcement and the finding that CR students who had mixed models were more likely to progress to a coherent model are in line with previous research. Wu, Anderson, and Miller (in press) have shown that students feel that CR helps them think, read, and write better. The findings of this study suggest that CR did indeed motivate them to think more and, in many cases, in successful ways. What is more, this occurred even when students were not explicitly told that a discussion was forthcoming.

CR discussions had an effect on the learning of students beyond the benefit of the actual discussions. The context of argumentative discussion alone is a factor in student learning. This departs from much of the previous research on discussion which investigates the effects of the discourse about a topic on learning of that topic. This study suggests that argumentative discussion can contribute to an overall classroom culture of argumentation with generalized effects that extend beyond the specific topics discussed.
Implications for Warm Conceptual Change Research

The results are consistent with a warm approach to conceptual change, because students who experienced CR discussions and began with mixed pre-test models had greater conceptual growth, even though they received no additional information. The fact that almost ninety percent of these students experienced conceptual growth cannot be attributed to being exposed to additional arguments or any additional intervention that confronted the students’ prior conceptions. Instead, this effect was created by introducing students to a new social discourse practice.

While further research is needed to explore the mechanisms behind the generalized effects of experiencing argumentative discourse, the fact that CR students increased their reading time suggests that the mechanism was related to increased processing as described in dual-process theories like the CRKM. However, what was not investigated in this particular study was if this new discourse context affected processing in the manner envisioned by the conceptual ecologies approach or as envisioned by the learning ecologies approach. To investigate this, it would need to be determined if students’ new models were not only different in the way that they modeled the Earth, but if they also took on different discourse and affective features consistent with the CR context. As discussed below this could be an interesting direction for future research.

Study Limitations and Areas for Future Research

The design of the study led to several limitations. Most importantly, the study design involved very short interventions. The CR treatment consisted of only four twenty minute discussions, and the announcement treatment was one sentence. Furthermore the text intervention consisted of a single episode of reading lasting less than ten minutes. These brief
interventions did cause change, but the change was incremental. This might be more properly categorized as conceptual growth rather than conceptual change, or it might be closer to Chi’s (2008) idea of belief change. To better investigate dual-process models of conceptual change in the future, methodologies will be needed to capture depth of processing during the more robust interventions needed to create radical conceptual change. A second related limitation was that the delayed post-test was not delayed enough to show any differences. For this reason the prediction that depth of processing leads to more persistent conceptual change was not testable. The strongest test of the theory would be under circumstances that created radical changes and in which those changes were tested after a long delay. This would require more robust interventions, larger sample sizes, and new methodologies.

In terms of reading times, the clause by clause method was successful, but did not provide enough detail to detect explicit signs in the reading time record of students using a dual-focus processing strategy. This might be detectable using eye-tracking equipment. First, a sample of participants who were instructed to use a dual-focus processing strategy would need to be analyzed. The patterns found in this group could then be compared to the patterns of participants asked to prepare for an argumentative discussion. The recent discovery that eye movements change when readers’ minds are not focused on the text (Reichle, Reineberg, & Schooler, 2010) suggests that such an approach might be successful.

As stated earlier, the SVT comprehension test had an unfortunate performance ceiling, because it was comprised of a small number of true and false questions. This could be remedied by a longer test with questions having a greater range of difficulty.

The self-assessment measure also had some limitations. The self-assessment measure consisted of a limited number of researcher made questions. Answers to the questions were not
very revealing about what students were actually thinking about while reading the text. Another research paradigm such as the “think aloud” method might provide more insight into what students think about while reading a text in the context of argumentative discussion. The think aloud approach was not chosen for this study because it makes reading time measurements impossible, and because it is labor intensive to analyze in conjunction with interview data. However, now that the reading time results have been established, a think aloud protocol could serve as a valuable follow-up study.

Finally, the analysis of the shape of the Earth interview data was a limitation of the study. Much of the rich information contained in the interviews went unanalyzed as the interview scoring method distilled each set of interviews into a binary outcome measure. A future analysis that made full use of the interviews might reveal further aspects of the conceptual growth process.

One possible direction for future analysis of the interviews would be to track changes in individual aspects of the shape of the Earth model (e.g. does gravity have a universal down direction). This would give a finer grained analysis of what exactly was changing between the first and second interviews and provide a more sensitive measure of conceptual growth. This might make it possible to statistically link the depth of processing of the text with later conceptual change possibly using structural equation modeling as well as qualitative methods.

Another promising direction for analysis would be judging changes between the interviews against information provided in the text. As illustrated in several of the interview answers displayed earlier in this dissertation, some students expressed new conclusions in the post-test interview that they attributed to the text, although the basis for the conclusions did not actually appear in the text. A further source of information on this topic is a yet unanalyzed post-
interview question in which students were explicitly asked if they thought they had changed their answers between interviews and if they thought the story influenced these changes. This type of analysis might enable an evaluation of how much the text influenced change by providing new information as compared to facilitating student’s thinking about their own ideas. Finally, an analysis could be undertaken to directly link discourse during the discussions with discourse during the interviews. The type of discourse used in CR discussions has been shown to influence students’ individual thinking as expressed in written essays (Reznitskaya et al., 2009). An analysis of the way in which students answer questions could demonstrate if similar carry over effects occur in shape of the Earth interviews and if these effects facilitate conceptual change.

**Educational Implications and Areas for Future Research**

The main education implication of this study is that argumentative classroom discussion can have a powerful impact beyond the specific topics of discussions by creating a context for thoughtful learning. The way students read a text was changed in measurable and important ways by the fact that it was being read as part of an argumentative classroom practice This should encourage teachers to take their classroom discourse practices seriously. The discourse in a classroom is not just a supplement to reading; it becomes part of the reading process itself. The nature of classroom discourse is especially important during science instruction in which students need to grapple with complex ideas. This study suggests that the effect of science texts on conceptual growth is significantly affected by the way that the texts are incorporated into the overall classroom discourse. While future research is required, this might be also true of many other types of instruction including labs and seat work.

However, not all the results of this study might be seen as positive by science teachers. This study suggests that reading for comprehension and reading for conceptual change are
different tasks, and they might have trade-off. To read for conceptual change a reader needs to be thinking about their own beliefs in addition to the text. Unless a person is rereading sections or going very slowly there may be compromises made about allocating cognitive resources.

The possibility that purposes for reading compete with on another has potential implications for the science teacher. In particular, it emphasizes the importance of teachers having a clear idea of their educational objectives and structuring the entire lesson, including reading tasks, according to those objectives. If a teacher is trying to get students to create and develop well-reasoned personally held beliefs about science topics, this research gives support to the idea of embedding reading and other science activities in an argumentative discussion context. This is in contrast to delivering content in a test-taking context or in an engineering-design context. However, if the objective of the teacher is to get students to learn and remember the details of scientific content, then the argumentative context might not be the most effective.

Science teachers often have many goals at once. They want students to memorize facts, acquire concepts, master scientific process skills, learn skills of inquiry and argument, learn about the nature of the scientific process, apply their knowledge to real world problems, and adopt personally held beliefs consistent with scientifically accepted theories. Any of these goals may be legitimate, but this research suggests that there are different requirements for reaching various goals, so that it may be difficult to address all simultaneously. Perhaps, it is necessary to revisit material multiple times in different ways in order for all these goals to be met. When it is the main goal of the teacher to get students to truly reflect on their own beliefs and to form cogent ideas about science topics that will last a lifetime, learning in the context of a social argumentative discourse approach like CR is a promising method.
Table 1.

Demographic Information for the Participating District and Schools

<table>
<thead>
<tr>
<th></th>
<th>Caucasian</th>
<th>African-American</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Receiving free or reduced lunch</th>
<th>English language learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>District</td>
<td>72.3</td>
<td>12.9</td>
<td>4.8</td>
<td>4.5</td>
<td>23.5</td>
<td>1.4</td>
</tr>
<tr>
<td>School 1</td>
<td>54.5</td>
<td>2.8</td>
<td>32.8</td>
<td>4.8</td>
<td>55.3</td>
<td>17.7</td>
</tr>
<tr>
<td>School 2</td>
<td>62.6</td>
<td>16.5</td>
<td>4.3</td>
<td>8.7</td>
<td>34.6</td>
<td>5.1</td>
</tr>
<tr>
<td>School 3</td>
<td>52.8</td>
<td>28.2</td>
<td>3.8</td>
<td>2.4</td>
<td>44.3</td>
<td>1.4</td>
</tr>
<tr>
<td>School 4</td>
<td>63.8</td>
<td>8.4</td>
<td>2.7</td>
<td>19.6</td>
<td>11.7</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Note. All values represent percentage of the total number of students.
Table 2.

*Experiment Timeline*

<table>
<thead>
<tr>
<th>Week 1 - 2 meetings</th>
<th>Pretests –</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Reading comprehension</td>
</tr>
<tr>
<td></td>
<td>2. Reading Fluency</td>
</tr>
<tr>
<td></td>
<td>3. Spatial Reasoning</td>
</tr>
<tr>
<td></td>
<td>4. Need for Cognition Questionnaire</td>
</tr>
<tr>
<td></td>
<td>5. Self-assessment of opinion confidence.</td>
</tr>
<tr>
<td></td>
<td>6. Shape of the Earth Interview.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 2 and 3</th>
<th>One half of the classrooms conducted Four CR Discussions – Videotaped</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 meetings each week for CR,</td>
<td>One half of classrooms received regular language art curriculum. -One session videotaped</td>
</tr>
<tr>
<td>1 meeting for Regular</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 4 - 1 Meeting</th>
<th>1. Students read the story “What Shape is the Earth?” on the computer. Half of the students in each classroom were first instructed that a discussion would follow the reading.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. After the reading, all students completed the SVT comprehension test and the self-assessment of depth of processing.</td>
</tr>
<tr>
<td></td>
<td>3. One half of students then received the Shape of the Earth Interview, and one-half received the epistemological interview.</td>
</tr>
</tbody>
</table>

| Week 5 - 1 Meeting | Delayed Posttest – Each student received whichever interview they had not received in the previous week. |

| Week 6 - 1 Meeting | Shape of the Earth discussion followed by a full class debriefing – Videotaped |
Table 3

Means, Standard Deviations, and T-test Comparisons For Pre-test Measures by Treatment Group

<table>
<thead>
<tr>
<th></th>
<th>Regular M</th>
<th>SD</th>
<th>CR M</th>
<th>SD</th>
<th>t – value</th>
<th>P - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cogat</td>
<td>8.46</td>
<td>3.78</td>
<td>9.02</td>
<td>4.11</td>
<td>-0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>NFC</td>
<td>46.02</td>
<td>8.79</td>
<td>44.33</td>
<td>9.15</td>
<td>0.95</td>
<td>0.35</td>
</tr>
<tr>
<td>Fluency</td>
<td>65.56</td>
<td>15.09</td>
<td>65.54</td>
<td>13.73</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Sat-10</td>
<td><strong>49.22</strong></td>
<td><strong>17.23</strong></td>
<td><strong>57.71</strong></td>
<td><strong>20.37</strong></td>
<td><strong>-2.26</strong></td>
<td><strong>0.026</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Unannounced M</th>
<th>SD</th>
<th>Announced M</th>
<th>SD</th>
<th>t – value</th>
<th>P - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cogat</td>
<td>9.20</td>
<td>3.96</td>
<td>8.40</td>
<td>3.94</td>
<td>1.01</td>
<td>0.31</td>
</tr>
<tr>
<td>NFC</td>
<td>45.53</td>
<td>9.15</td>
<td>44.81</td>
<td>8.90</td>
<td>0.4</td>
<td>0.69</td>
</tr>
<tr>
<td>Fluency</td>
<td>66.87</td>
<td>14.47</td>
<td>64.51</td>
<td>14.24</td>
<td>0.82</td>
<td>0.41</td>
</tr>
<tr>
<td>Sat-10</td>
<td>53.05</td>
<td>19.11</td>
<td>54.24</td>
<td>19.70</td>
<td>-0.31</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Note: bold comparisons are different at the .05 level.
Table 4.

*Intercorrelations Between Clause Level Reading Time Variables*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables</td>
<td>-0.057**</td>
<td>-0.142**</td>
<td>0.643**</td>
<td>-0.008</td>
<td>0.280**</td>
<td>-0.050**</td>
<td>-0.180**</td>
<td>0.117**</td>
<td>-0.028*</td>
</tr>
<tr>
<td>Frequency</td>
<td>--</td>
<td>0.563**</td>
<td>0.116**</td>
<td>0.010</td>
<td>-0.288**</td>
<td>0.164**</td>
<td>0.473**</td>
<td>-0.218**</td>
<td>0.165**</td>
</tr>
<tr>
<td>Content Freq.</td>
<td>--</td>
<td>0.141**</td>
<td>0.087**</td>
<td>-0.246**</td>
<td>0.161**</td>
<td>0.375**</td>
<td>-0.156**</td>
<td>0.150</td>
<td></td>
</tr>
<tr>
<td>Propositions</td>
<td>--</td>
<td>-0.019*</td>
<td>0.154**</td>
<td>0.008</td>
<td>-0.012</td>
<td>0.121**</td>
<td>-0.047**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serial Order</td>
<td>--</td>
<td>-0.008</td>
<td>-0.007</td>
<td>0.009</td>
<td>-0.014</td>
<td>0.029*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Arg.</td>
<td>--</td>
<td>-0.087**</td>
<td>-0.346**</td>
<td>0.209**</td>
<td>-0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg. Wrap-Up</td>
<td>--</td>
<td>0.210**</td>
<td>-0.111**</td>
<td>-0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg. Order</td>
<td>--</td>
<td>-0.033**</td>
<td>0.408**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After Picture</td>
<td>--</td>
<td>0.070**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01
Table 5

*Reading Time Mixed Effects Model Results*

| Effect                          | Standard Estimate | Error    | DF    | t Value | Pr > |t| |
|--------------------------------|-------------------|----------|-------|---------|-------|---|
| Intercept                      | 8.220             | 0.113    | 103   | 72.85   | <.001 | |
| Propositions                  | 0.134             | 0.00666  | 12,000| 20.05   | <.001 | |
| Serial Order                   | -0.00247          | 0.000418 | 12,000| -5.92   | <.001 | |
| New Argument                   | 0.0913            | 0.0247   | 12,000| 3.69    | <.001 | |
| Fluency                        | -0.00152          | 0.000176 | 12,000| -8.62   | <.001 | |
| Discussion                     | 0.165             | 0.0463   | 12,000| 3.57    | <.001 | |
| Announcement                   | 0.0616            | 0.0568   | 12,000| 1.08    | 0.278 | |
| Propositions x Discussion      | 0.0163            | 0.00586  | 12,000| 2.78    | 0.006 | |
| Propositions x Announcement    | -0.0108           | 0.00574  | 12,000| -1.88   | 0.060 | |
| New Argument x Discussion      | 0.0255            | 0.0222   | 12,000| 1.15    | 0.250 | |
| New Argument x Announcement    | -0.0136           | 0.0217   | 12,000| -0.63   | 0.531 | |
| Serial Order x Discussion      | 0.00026           | 0.000252 | 12,000| 1.03    | 0.301 | |
| Serial Order x Announcement    | 0.00007           | 0.000253 | 12,000| 0.28    | 0.783 | |
Table 6

*Logistic Regression Results Predicting Students who Skip More Than 10% of Clauses*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.081</td>
<td>0.006</td>
<td>189.542</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gender</td>
<td>-1.546</td>
<td>0.069</td>
<td>502.331</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fluency</td>
<td>-0.0005</td>
<td>0.0003</td>
<td>2.072</td>
<td>0.150</td>
</tr>
<tr>
<td>CogAt</td>
<td>0.060</td>
<td>0.010</td>
<td>35.492</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Need for Cognition</td>
<td>0.051</td>
<td>0.004</td>
<td>142.923</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stanford-10</td>
<td>-0.009</td>
<td>0.003</td>
<td>11.975</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sphere/Non-sphere Model</td>
<td>-15.581</td>
<td>180.8</td>
<td>0.0074</td>
<td>0.931</td>
</tr>
<tr>
<td>Strict/Non-strict Model</td>
<td>1.206</td>
<td>0.119</td>
<td>103.029</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hollow/Non-hollow Model</td>
<td>-16.157</td>
<td>180.8</td>
<td>0.008</td>
<td>0.929</td>
</tr>
<tr>
<td>Mixed/Coherent Model</td>
<td>7.868</td>
<td>90.420</td>
<td>0.0076</td>
<td>0.931</td>
</tr>
<tr>
<td>Discussion</td>
<td>0.403</td>
<td>0.037</td>
<td>120.915</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Announcement</td>
<td>-0.352</td>
<td>0.037</td>
<td>89.943</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Discussion x Announcement</td>
<td>-0.356</td>
<td>0.036</td>
<td>98.928</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table 7

Sentence Verification Test Regression Analysis Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford 10 (SAT-10)</td>
<td>1</td>
<td>0.004</td>
<td>0.002</td>
<td>7.31</td>
<td>0.007</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>-0.069</td>
<td>0.086</td>
<td>0.64</td>
<td>0.422</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>0.239</td>
<td>0.084</td>
<td>8.02</td>
<td>0.005</td>
</tr>
<tr>
<td>SAT-10 x Discussion</td>
<td>1</td>
<td>0.001</td>
<td>0.002</td>
<td>0.71</td>
<td>0.398</td>
</tr>
<tr>
<td>SAT-10 x Announcement</td>
<td>1</td>
<td>-0.005</td>
<td>0.002</td>
<td>9.53</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Table 8

*Descriptive Statistics for Self-Assessment of Depth of Processing*

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) I think I understood this story pretty well.</td>
<td>3.950</td>
<td>1.161</td>
<td>5</td>
</tr>
<tr>
<td>2) The story was interesting to read.</td>
<td>3.752</td>
<td>1.276</td>
<td>5</td>
</tr>
<tr>
<td>3) I put a lot of effort into reading and understanding this story.</td>
<td>3.772</td>
<td>1.272</td>
<td>5</td>
</tr>
<tr>
<td>4) This story reminded me of other things that I've already read and heard about.</td>
<td>3.574</td>
<td>1.431</td>
<td>5</td>
</tr>
<tr>
<td>5) When I was reading this story, I sometimes stopped reading to think a little while before I kept reading.</td>
<td>3.238</td>
<td>1.379</td>
<td>5</td>
</tr>
<tr>
<td>6) Reading this story made me think a lot about the shape of the Earth.</td>
<td>3.832</td>
<td>1.273</td>
<td>5</td>
</tr>
<tr>
<td>7) When I was reading this story, I asked myself if I really believed what I was reading.</td>
<td>3.119</td>
<td>1.551</td>
<td>5</td>
</tr>
<tr>
<td>8) Knowing about the shape of the Earth is important to me.</td>
<td>3.743</td>
<td>1.262</td>
<td>5</td>
</tr>
<tr>
<td>9) When I was reading this story, I tried to picture in my mind the things I was reading about.</td>
<td>3.861</td>
<td>1.209</td>
<td>5</td>
</tr>
<tr>
<td>10) When I read about Jill and Maria and they did not agree with each other, it made me confused.</td>
<td>2.545</td>
<td>1.439</td>
<td>5</td>
</tr>
<tr>
<td>11) When I read about Jill and Maria and they did not agree with each other, I chose one of them to believe.</td>
<td>3.594</td>
<td>1.498</td>
<td>5</td>
</tr>
<tr>
<td>12) When I read this story, I thought about what I might say during my ____ discussion.</td>
<td>3.594</td>
<td>1.282</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 9

Results of the Principal Components Analysis of the Self-assessment Measure of Student’s Depth of Processing

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.237</td>
<td>0.899</td>
<td>0.672</td>
<td>0.672</td>
</tr>
<tr>
<td>2</td>
<td>1.338</td>
<td>1.027</td>
<td>0.402</td>
<td>1.073</td>
</tr>
<tr>
<td>3</td>
<td>0.311</td>
<td>0.154</td>
<td>0.093</td>
<td>1.167</td>
</tr>
<tr>
<td>4</td>
<td>0.158</td>
<td>0.011</td>
<td>0.047</td>
<td>1.214</td>
</tr>
<tr>
<td>5</td>
<td>0.147</td>
<td>0.056</td>
<td>0.044</td>
<td>1.258</td>
</tr>
<tr>
<td>6</td>
<td>0.090</td>
<td>0.080</td>
<td>0.027</td>
<td>1.285</td>
</tr>
<tr>
<td>7</td>
<td>0.010</td>
<td>0.056</td>
<td>0.003</td>
<td>1.288</td>
</tr>
<tr>
<td>8</td>
<td>-0.046</td>
<td>0.043</td>
<td>-0.014</td>
<td>1.275</td>
</tr>
<tr>
<td>9</td>
<td>-0.089</td>
<td>0.139</td>
<td>-0.027</td>
<td>1.248</td>
</tr>
<tr>
<td>10</td>
<td>-0.228</td>
<td>0.031</td>
<td>-0.068</td>
<td>1.179</td>
</tr>
<tr>
<td>11</td>
<td>-0.259</td>
<td>0.0805</td>
<td>-0.078</td>
<td>1.102</td>
</tr>
<tr>
<td>12</td>
<td>-0.339</td>
<td>-0.102</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note: Eigenvalues of the Reduced Correlation Matrix:

Total = 3.33178309  Average = 0.27764859
Table 10

*Factor Loadings after Promax Rotation*

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.437</td>
<td>-0.424</td>
</tr>
<tr>
<td>2</td>
<td>0.636</td>
<td>0.156</td>
</tr>
<tr>
<td>3</td>
<td>0.271</td>
<td>0.591</td>
</tr>
<tr>
<td>4</td>
<td>0.255</td>
<td>-0.329</td>
</tr>
<tr>
<td>5</td>
<td>0.146</td>
<td>0.431</td>
</tr>
<tr>
<td>6</td>
<td>0.372</td>
<td>0.372</td>
</tr>
<tr>
<td>7</td>
<td>0.387</td>
<td>0.447</td>
</tr>
<tr>
<td>8</td>
<td>0.545</td>
<td>0.186</td>
</tr>
<tr>
<td>9</td>
<td>0.530</td>
<td>0.192</td>
</tr>
<tr>
<td>10</td>
<td>0.027</td>
<td>0.594</td>
</tr>
<tr>
<td>11</td>
<td>0.187</td>
<td>-0.016</td>
</tr>
<tr>
<td>12</td>
<td>0.640</td>
<td>0.118</td>
</tr>
</tbody>
</table>
Table 11

Regression Results for the Total Score of the Self-Assessment of Student’s Depth of Processing Survey

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for Cognition</td>
<td>1</td>
<td>0.211</td>
<td>0.073</td>
<td>8.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Spherical/Non-Spherical Model</td>
<td>1</td>
<td>-5.999</td>
<td>1.341</td>
<td>20.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>-1.535</td>
<td>0.666</td>
<td>5.31</td>
<td>0.021</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>-0.932</td>
<td>0.656</td>
<td>2.01</td>
<td>0.156</td>
</tr>
</tbody>
</table>
Table 12

Regression Results for Factor One of the Self-Assessment of Student’s Depth of Processing Survey

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for Cognition</td>
<td>1</td>
<td>0.172</td>
<td>0.056</td>
<td>9.31</td>
<td>0.002</td>
</tr>
<tr>
<td>Spherical/Non-Spherical Model</td>
<td>1</td>
<td>-3.155</td>
<td>1.038</td>
<td>9.25</td>
<td>0.002</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>-0.874</td>
<td>0.516</td>
<td>2.88</td>
<td>0.090</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>-0.487</td>
<td>0.508</td>
<td>0.92</td>
<td>0.337</td>
</tr>
</tbody>
</table>
Table 13

*Regression Results for Factor Two of the Self-Assessment of Student’s Depth of Processing Survey*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford 10</td>
<td>1</td>
<td>-0.078</td>
<td>0.029</td>
<td>7.19</td>
<td>0.007</td>
</tr>
<tr>
<td>Spherical/ Non-Spherical Model</td>
<td>1</td>
<td>-3.177</td>
<td>1.074</td>
<td>8.74</td>
<td>0.003</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>-0.431</td>
<td>0.511</td>
<td>0.71</td>
<td>0.399</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>-0.828</td>
<td>0.492</td>
<td>2.83</td>
<td>0.092</td>
</tr>
</tbody>
</table>
Table 14

Regression Results for Question 11 of the Self-Assessment of Student’s Depth of Processing Survey

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need for Cognition</td>
<td>1</td>
<td>0.040</td>
<td>0.017</td>
<td>6.82</td>
<td>0.009</td>
</tr>
<tr>
<td>Spherical/Non-Spherical Model</td>
<td>1</td>
<td>-0.471</td>
<td>0.309</td>
<td>2.32</td>
<td>0.127</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>0.157</td>
<td>0.153</td>
<td>1.04</td>
<td>0.307</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>-0.157</td>
<td>0.151</td>
<td>1.08</td>
<td>0.299</td>
</tr>
</tbody>
</table>
Table 15

**Means for Items on the Self-Assessment Survey Divided by Discussion**

**Treatment**

<table>
<thead>
<tr>
<th>Question</th>
<th>Regular</th>
<th>CR</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.15</td>
<td>3.77</td>
<td>1.161</td>
</tr>
<tr>
<td>2</td>
<td>3.83</td>
<td>3.68</td>
<td>1.276</td>
</tr>
<tr>
<td>3</td>
<td>4.00</td>
<td>3.57</td>
<td>1.272</td>
</tr>
<tr>
<td>4</td>
<td>3.77</td>
<td>3.40</td>
<td>1.431</td>
</tr>
<tr>
<td>5</td>
<td>3.44</td>
<td>3.06</td>
<td>1.379</td>
</tr>
<tr>
<td>6</td>
<td>4.31</td>
<td>3.40</td>
<td>1.273</td>
</tr>
<tr>
<td>7</td>
<td>3.27</td>
<td>2.98</td>
<td>1.551</td>
</tr>
<tr>
<td>8</td>
<td>3.81</td>
<td>3.68</td>
<td>1.262</td>
</tr>
<tr>
<td>9</td>
<td><strong>3.85</strong></td>
<td><strong>3.87</strong></td>
<td><strong>1.209</strong></td>
</tr>
<tr>
<td>10</td>
<td>2.75</td>
<td>2.36</td>
<td>1.439</td>
</tr>
<tr>
<td>11</td>
<td><strong>3.50</strong></td>
<td><strong>3.68</strong></td>
<td><strong>1.498</strong></td>
</tr>
<tr>
<td>12</td>
<td>3.65</td>
<td>3.55</td>
<td>1.282</td>
</tr>
</tbody>
</table>

Note: Standard Deviations are calculated across all participants. Bold items are questions for which CR students did not rate themselves lower than students in the regular instruction classrooms.
Table 16

*Results of the Shape of the Earth Interview*

<table>
<thead>
<tr>
<th></th>
<th>sphere</th>
<th>sphere</th>
<th>hollow</th>
<th>hollow</th>
<th>dual</th>
<th>flattened</th>
<th>disc</th>
<th>mixed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>18</td>
<td>23</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>loose</td>
<td>44</td>
<td>15</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96</td>
</tr>
</tbody>
</table>

Pre  102
Post 96
Table 17

*Results of the Shape of the Earth Interview Collapsed Across Loose and Strict Models*

<table>
<thead>
<tr>
<th></th>
<th>sphere</th>
<th>hollow</th>
<th>dual</th>
<th>flattened</th>
<th>disc</th>
<th>Mixed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>41</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>38</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>40.2%</td>
<td>14.7%</td>
<td>2.0%</td>
<td>1.0%</td>
<td>2.0%</td>
<td>37.3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>38</td>
<td>102</td>
</tr>
<tr>
<td>Post</td>
<td>59</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>19</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>61.5%</td>
<td>13.5%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>2.1%</td>
<td>18.6%</td>
<td></td>
</tr>
</tbody>
</table>
Table 18

Summary of Results of the Shape of the Earth Interview Reported in Vosniadou and Brewer (1992)

<table>
<thead>
<tr>
<th></th>
<th>sphere</th>
<th>hollow</th>
<th>dual</th>
<th>flattened</th>
<th>disc</th>
<th>Mixed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd Grade</td>
<td>8  40%</td>
<td>4  20%</td>
<td>2  10%</td>
<td>3  15%</td>
<td>1  5%</td>
<td>2  10%</td>
<td>20</td>
</tr>
<tr>
<td>5th Grade</td>
<td>12  60%</td>
<td>6  30%</td>
<td>0  0%</td>
<td>0  0%</td>
<td>0  0%</td>
<td>2  10%</td>
<td>20</td>
</tr>
</tbody>
</table>
Table 19

Results of the Conceptual Change Analysis for All Participants

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td>16</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>hollow strict</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>hollow loose</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on left side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 20

*Results of the Conceptual Change Analysis for Students in Collaborative Reasoning Condition*

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td></td>
<td>10</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>hollow strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow loose</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on left side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 21

Results of the Conceptual Change Analysis for Students in the Regular Instruction Condition

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow loose</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>dual strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on right side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 22

*Results of the Conceptual Change Analysis for Students in the Announcement Condition*

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td></td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow strict</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on right side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 23

Results of the Conceptual Change Analysis for Students in the No Announcement Condition

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>hollow strict</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>hollow loose</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>flattened</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on right side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 24

**Results of the Conceptual Change Analysis for Students Receiving an Immediate Post-test**

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow strict</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow loose</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual strict</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on the right side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 25

*Results of the Conceptual Change Analysis for Students Receiving a Delayed Post-test*

<table>
<thead>
<tr>
<th></th>
<th>strict</th>
<th>loose</th>
<th>hollow</th>
<th>hollow</th>
<th>flattened</th>
<th>flattened</th>
<th>dual strict</th>
<th>dual loose</th>
<th>disc strict</th>
<th>disc loose</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>strict</td>
<td>11</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loose</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hollow strict</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>hollow loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>flattened strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>flattened loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dual loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc strict</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disc loose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mixed</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Note. Pre-test scores are listed on right side of the table. Post-test scores are listed on the top of the table. Students who had no change can be seen along the diagonal, students who progress are seen below the diagonal (in bold), and students who regressed are seen above the diagonal (in italics).
Table 26

Results for Logistic Regression of Shape of the Earth Model Pre-test Post-test Improvement

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Abilities Test</td>
<td>1</td>
<td>-0.396</td>
<td>0.119</td>
<td>11.0506</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mixed/Coherent Model</td>
<td>1</td>
<td>-2.216</td>
<td>0.843</td>
<td>6.901</td>
<td>0.0086</td>
</tr>
<tr>
<td>Discussion</td>
<td>1</td>
<td>0.386</td>
<td>0.475</td>
<td>0.661</td>
<td>0.416</td>
</tr>
<tr>
<td>Announcement</td>
<td>1</td>
<td>0.311</td>
<td>0.348</td>
<td>0.799</td>
<td>0.371</td>
</tr>
<tr>
<td>Mixed/Coherent Model x Discussion</td>
<td>1</td>
<td>-1.531</td>
<td>0.718</td>
<td>4.552</td>
<td>0.0329</td>
</tr>
</tbody>
</table>
Figure 2. Diagram showing the Cognitive Reconstruction of Knowledge Model. Reprinted from “Reconceptualizing Change in the Cognitive Construction of Knowledge” by J. Dole and G. M Sinatra, 1998, *Educational Psychologist*, 33, p. 119.
Figure 4. Examples of the five most common types of pictures drawn by students who were categorized as having a spherical model of the Earth (See Appendix E for definitions of each category). The first row shows examples of pictures in category one, three, and five. The second row shows examples of pictures in category five and seven.
Figure 5. Example of the only type of drawing that was considered consistent with a hollow Earth model but not a spherical model.
Figure 6. Drawing of a student classified as having a dual Earth model.
Figure 7. Drawing of student categorized as having a disc-shaped model of the Earth.
Figure 8. Histograms and Q-Q plots for reading times. Top two panels show frequency histograms for reading times. The left histogram is plots raw reading times and the right histogram plots log transformed reading times. The two bottom panels show the Q-Q plots of the reading times. The left histogram is plots raw reading times and the right histogram plots log transformed reading times. The solid line in each graph shows the expected values for a normal distribution given the mean and standard deviation of the data set.
Figure 9. Scatterplot of students’ Stanford 10 Reading Comprehension scores and Need for Cognition Survey total score. The asterisks represent values of students who skipped less than 10% of the clauses. The triangles represent the values of students who skipped more than 10% of the clauses. The two lines represent regression lines independently fitted to each group. The solid line is the regression line for students who skipped less than 10% of the clauses. The dotted line is the regression line for students who skipped more than 10% of the clauses.
Figure 10. Histograms and Q-Q plots for SVT test results. Top two panels show frequency histograms for the SVT test. The left histogram is plots raw scores and the right histogram plots square root arcsine transformed scores. The two bottom panels show the Q-Q plots of the SVT test. The left histogram is plots raw scores and the right histogram plots square root arcsine transformed scores. The solid line in each graph shows the expected values for a normal distribution given the mean and standard deviation of the data.
Figure 11. Scatterplot of students’ Stanford 10 Reading Comprehension scores and total score on the SVT test. The plus signs represent values of students who received an announcement of a future discussion. The triangles represent the values of students who did not receive an announcement of a future discussion. The two lines represent regression lines independently fitted to each group. The solid line is the regression line for students who did not receive an announcement. The dashed line is the regression line for students who received an announcement.
Figure 12. Scree Plot showing results of the principal components analysis of the self-assessment measure of student’s depth of processing.
Figure 13. Bar graph showing the percentage of students whose models progressed from their pre-test to their post-test model.
Figure 14. Drawings of students who showed a radical conceptual change. The drawing on the left is from the pre-test interview of a child who began with a dual-Earth model. The drawing on the right is from the post-test of the same child who now adopted a hollow Earth model.
References


Miller, B. (2009a). *A coat for Mr. Snowman*. Champaign, IL: Center for the Study of Reading.


Appendix A

Norms of discussion, norms of argument, and instructional moves for collaborative reasoning discussions

**Norms of Discussion**
- We speak freely without raising your hand.
- We do not talk when other people are talking.
- We talk among one another—not only to teacher.
- We encourage everyone to participate.
- We listen carefully to everyone’s ideas.
- We think about ideas and not people.

**Norms of Argument**
- We support our positions with reasons and evidence.
- We consider different sides of an issue.
- We challenge others’ ideas.

**Nine Instructional Moves**
1. Prompting
2. Thinking out loud
3. Asking for clarification
4. Challenging
5. Step in and Remind
6. Encouraging
7. Fostering independence
8. Summing up and refocusing
9. Debriefing
## Appendix B

### The Reading Fluency Sentence Verification Test

<table>
<thead>
<tr>
<th></th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lions like to eat grass.</td>
</tr>
<tr>
<td>2</td>
<td>An ant is smaller than an elephant.</td>
</tr>
<tr>
<td>3</td>
<td>A car is much longer than a train.</td>
</tr>
<tr>
<td>4</td>
<td>Cats can catch mice.</td>
</tr>
<tr>
<td>5</td>
<td>Bicycles go faster than airplanes.</td>
</tr>
<tr>
<td>6</td>
<td>The sun gives us light and heat.</td>
</tr>
<tr>
<td>7</td>
<td>A baseball is bigger than a basketball.</td>
</tr>
<tr>
<td>8</td>
<td>There are people living on the moon.</td>
</tr>
<tr>
<td>9</td>
<td>A mushroom is a kind of room.</td>
</tr>
<tr>
<td>10</td>
<td>It never snows in Alaska.</td>
</tr>
<tr>
<td>11</td>
<td>We can find a lot of books in the library.</td>
</tr>
<tr>
<td>12</td>
<td>Babies can walk as soon as they are born.</td>
</tr>
<tr>
<td>13</td>
<td>Turtles have long ears.</td>
</tr>
<tr>
<td>14</td>
<td>The flu is a kind of sickness.</td>
</tr>
<tr>
<td>15</td>
<td>All the flowers in the world are red.</td>
</tr>
<tr>
<td>16</td>
<td>A square has four sides.</td>
</tr>
<tr>
<td>17</td>
<td>A zebra has no strips.</td>
</tr>
<tr>
<td>18</td>
<td>A bell makes a ringing sound.</td>
</tr>
<tr>
<td>19</td>
<td>Books are made of paper.</td>
</tr>
<tr>
<td>20</td>
<td>A pencil is used for eating food.</td>
</tr>
<tr>
<td>21</td>
<td>Roses bloom in the winter.</td>
</tr>
<tr>
<td>22</td>
<td>Fish need water to live.</td>
</tr>
<tr>
<td>23</td>
<td>We breathe air.</td>
</tr>
<tr>
<td>24</td>
<td>People usually wear shoes on their hands.</td>
</tr>
<tr>
<td>25</td>
<td>The milk we drink comes from horses.</td>
</tr>
<tr>
<td>26</td>
<td>Baby dogs are called bunnies.</td>
</tr>
<tr>
<td>27</td>
<td>Exercise is good for your health.</td>
</tr>
<tr>
<td>28</td>
<td>Some people put salt on their food.</td>
</tr>
</tbody>
</table>
Eggs are hard as rocks.
We can see many animals in the zoo.
People use hands to walk and feet to eat.
Babies wear diapers.
You see steam when water boils.
We can tell the time from a chair.
Tigers are afraid of sheep.
Ice is frozen water.
Tea is made in a television.
We can get wet in the rain.
Snow is black and white.
There are 100 states in the U.S.
There are seven days in a week.
The sky is blue when it is sunny.
Many kinds of fish live in the ocean.
Elephants can fly.
Eating a lot of candy is good for your teeth.
The sun sets in the morning.
Your mother’s mother is your grandmother.
Bears sleep all winter.
Plants and animals need water to live.
Telephones are vegetables.
A raincoat is most useful when it is sunny.
When you look into the mirror, you see yourself in it.
At school we are taught to read and write.
Basketball is a team sport.
Peaches, plums, and apples are fruits.
You can buy food from a supermarket.
We should not be mean to our friends.
Christmas comes before Thanksgiving.
People have four legs.
60. We go to elementary school at the age of 80.
61. You can find camels and cactus in the desert.
62. We need to wear more clothes when it is cold.
63. A pitcher throws the ball during a football game.
64. The eyebrows are below the eyelashes.
65. Breakfast comes before dinner.
66. A violin is a string instrument.
67. Your shadow follows you wherever you go.
68. A triangle has five sides.
69. Hens lay eggs, and monkeys lay eggs, too.
70. An air conditioner can make us feel cooler.
71. There are twelve months in one year.
72. School buses are usually black.
73. A pizza is usually in a round shape.
74. A scissor is used to hold water.
75. Pandas are from China.
76. The sun looks small because it is far away.
77. The U. S. flag has stars and stripes.
78. A lid goes on top of a box.
79. The Moon goes around the Earth.
80. A cork can float in water.
81. Kangaroos carry their baby in their pouch.
82. An elevator can take you up and down.
83. Seventh grade students are elementary school students.
84. Cars can cross the street when the green light is on.
85. There are 20 minutes in an hour, and 80 hours in a day.
86. The earth is smaller than the moon.
87. Most computers are made of wood.
88. A computer needs electricity to run.
89. You need to peel the skin before eating an orange or banana.
90. Rugs are used to cover the floor.
91. A dime is worth more than a quarter.
92. Children are older than adults.
93. Most people have fifteen fingers on each hand.
94. Owls are usually awake during the night.
95. Your father’s brother is your uncle.
96. Letters are sent in the mail.
97. A five page story is longer than a seven page story.
98. Juice is a drink made from fruit.
99. Your hair will grow very long if you don’t cut it for a long time.
100. Cars are supposed to stop when they come to a stop sign.
101. George Washington was the first president of the United States.
102. Abraham Lincoln was the first president of Spain.
103. Christopher Columbus discovered France.
104. In Germany, most people speak Spanish.
105. The ancient Egyptians built pyramids.
106. A ladder is a kind of musical instrument.
107. A piano is a kind of musical instrument.
108. A hammer is used to put nails into wood.
109. We can see stars in the sky during a sunny day.
110. It is colder in the summer than it is in the winter.
Appendix C

The Need for Cognition Survey (modified for children)

How true are these statements to you? Fill in the circle that best describes you.

1. I like hard problems.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

2. I like to be the person who gets to make big decisions.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

3. Thinking is not my idea of fun.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

4. I do not like to work on problems that I know will be very difficult to figure out.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

5. When I am allowed, I choose work that will not make me think a lot.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

6. I like to do problems that take a long time to get the answer.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

7. I only think as hard as I have to.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

8. I like one big project more than many small projects.
   
   Not at all true  somewhat true  very true
   1  2  3  4  5

9. I like doing things that I can do quickly once I have learned them.
10. I like competitions where you can win if you are a good thinker.

11. I like to think of new ways of doing things and solving problems.

12. I like puzzles.

13. If something works, I don’t care why it works.

14. I like questions that are easy to answer.
Appendix D

Shape of the Earth interview questions and protocol

1) What is the shape of the earth?

2) How do you know the Earth is ______________? (Use the student’s word for the shape of the Earth)

3) Which direction do we look to see the earth? (If the child is confused, you can gesture by pointing in different directions)

4) Please draw a picture of the earth? (Give the child the clipboard and pencil.)

5) Now on this drawing show me where the moon and stars go. Please draw the moon and stars on your picture.

6) Now show me where the sky is on your picture. Please draw the sky on your picture. (if they are confused ask them to draw clouds).

7) Show me where the people live. (Child may draw or point. Describe any gestures)

8) (a) Here is a picture of a house. This house is on the earth, isn’t it?
(b) How come the earth is flat here, but before you said it was __________?

9) If you walked and walked for many days in a straight line, where would you end up?
(If the student asks about water, say that when you get to the water you would get in a boat a sail across in a straight line)

10) Does the Earth have an end or edge?

IF NO, ASK: Why not?

IF YES ASK: Could someone fall off the end or edge of the Earth? Where would they fall?

11) (a) Now I want you to show me where Bloomington/Normal is on your picture.
(b) Now I want you to show me where China is on your picture?
(c) Please draw a person in each place. (Write a B above the Bloomington person and a C above the Chinese person. Stick figures are fine.)

12) Now tell me what is down here below the earth.
(point/motion to an area underneath the Earth [from the child’s perspective] in their drawing).

13) Does anyone live, here, on the bottom of the Earth? (point to the same place underneath the Earth in their drawing?)
IF NO, ASK: Why not?
IF YES, ASK: Why don’t they fall off?

14) Is there anything more you would like to say about the Earth?

Note: Hold up the card as you read the question.

1) When you think about the shape of the Earth are you: very confused, confused, certain, or very certain

2) When you try to picture the Earth in your mind is it – very hard, hard, easy, or very easy

3) How sure are you that you have the correct idea about the shape of the Earth?
Very unsure, unsure, sure, or very sure

4) How many other people do you think agree with you about the shape of the Earth?
Very few people, a few people, a lot of people, almost all people

Is there anything more you would like to add?

Do you have any questions?
Time = 10 – 15 minutes per person

Materials - A Record Sheet for each student.
A digital recorder.
A clipboard with a blank sheet of paper
A sharpened pencil with an eraser (for the student).
A copy of the house drawing.
A copy of the confidence rating cards.
Paper Clips Stapler

Setting Up –

1) Find a relatively quiet place to sit (either floor or table)
2) Write down your name, the teacher’s name, and the number written on the back of the recorder.
3) Turn on the digital tape recorder by sliding down and holding the power switch in the down position for 2 seconds. It should say Folder A.
4) Call the name of the first student and direct them to where you are sitting.

Getting started –

1) Introduce yourself.
2) Write down the student’s name
3) Ask the student what month they were born. Ask what year they were born. Write down the student’s birth month and year on the Record Sheet.
4) Write the student’s name on the blank sheet of paper on the clipboard

Explain what the student will be doing and the purpose of the interview
I am interested in what children think about the Earth. I am going to ask you some questions and I’m going to ask you to make a drawing. I’m going to tape record us so that I can remember what you said. Do you have any questions? Are you ready to start?

Interview Tips –

This interview is semi-structured. That means you need to follow the questions, but you may prompt for more information or ask additional questions to clarify or skip questions that the child has already answered. However, we need to avoid suggesting any answers to the student. If a student is giving very short answers or is not explaining things completely, you should start by using gestures or simple questions. For example:

- Silence – (10 seconds)
- nodding your head
- “uh-huh”
- “Is there anything else you would like to say?”
- “Can you tell me more?”

Only resort to more explicit questions if these do not work. Even then try to use the student’s own words. For example:

- “What did you mean by ______?”
- “Can you show me on your drawing.”

Your job is to understand what the student thinks. Don’t forget they might not have an opinion on some questions. “I don’t know” is a viable answer. Ask additional questions only to clarify until you feel you understand the student’s intentions. Also remember that gestures and drawings are viable answers as well as spoken words.
Start the interview

1) PRESS REC and place the recorder on the table or the floor near the student.
2) SAY: *I need you to start by saying your first and last name for the recorder.* And have student say their name.
3) Continue through questions on the Record Sheet

Ending the interview

1) Press STOP on the recorder.
2) Write down the file directory (should be A) and file number on the Record Sheet (the numbers should be consecutive).
3) Thank the student.
4) Ask the student to send out the next student on your list.
5) Make any important notes. Write any notes on the picture needed to interpret ambiguous markings. Paper clip the record sheet and drawing sheet.

IF IT IS THE LAST STUDENT: turn off the recorder by sliding and holding the power switch for 2 seconds.
Appendix E

Shape of the Earth interview scoring rubric

The number of students giving each answer on the pre and post test is indicated in parenthesis following the answer. If no student gave the answer then the parenthesis were omitted. Students were allowed to give more than one answer for some questions. In the model matching section, the answers that were consistent with each model are listed followed by the answers that are acceptable deviations from the expected model consistent answers.

Q1  What is the Shape of the Earth?

(0) Don’t know   (1) round (13, 43)   (2) sphere/round like a ball (61, 53)
(3) oval (1, 0)   (4) circle (34, 19)

Model Matching   Sphere – 1, 2, 3   acceptable deviation – 4

Hollow – 1, 2, 3, 4

Flattened – 1, 3, 4

Dual – 1, 2, 3, 4

Disc – 1, 3, 4

Q2  How do you know the Earth is round?

(0) Don’t know   (1) Books   (2) TV/movies    (3) pictures and maps
(4) globe      (5) class/teacher   (6) mom/dad, friend other people
(7) from space program   (8) from Columbus   (9) From the story
(10) no explanation (because its round like a ball/ people live there etc.)
(11) If you go on you come back to where you started/ go round/ don’t fall

(12) N/A   (13) Other   (13) Missing data/incomprehensible

Not used for model matching or comprehensiveness
Q3  Which direction do we look to see the earth?

(0) Don’t know (4, 1)    (1) Down (22, 29)    (2) sideways (6, 6)
(3) everywhere/all around (20, 29)    (4) in the back/forward/straight ahead (6, 10)
(5) up (37, 22)    (6) North (13, 10)    (7) other (6, 5)    (8) Missing (0, 1)

(20) Uninterruptable (0, 1)

Model Matching  
Sphere  - 1,2,3,4 acceptable deviation – 5
Hollow – 1, 2, 3, 4, 5, 6
Flattened – 1,2,3,4 acceptable deviation – 5
Dual – 5
Disc  - 1,2,3 acceptable deviation – 5,6

Q4  Can you draw me a picture of the Earth?

(1) circle (sphere) (110, 117)    (2) oval
(3) flat line but changes to circle after questioning    (4) circle within square frame
(5) rectangle/straight line (1, 0)

Model Matching  
Sphere – 1, 2
Hollow – 1, 2
Flattened – 1, 2
Dual – 1, 2
Disc – 1, 2, 3
Q5 Now on this drawing show me where the moon and stars go.

Q6 Now show me where the sky is on your picture.

Model Matching

- Sphere – 1, 2, 3, 4, 5, 6, 7
- Hollow - 1, 2, 3, 4, 5, 6, 7, 8
- Flattened – 1, 2, 3, 4, 5
- Dual – 1, 2, 3, 4, 7, 9, 10
- Disc – 1, 2, 8

Q7 Show me where the people live.
Q8 (a) Here is a picture of a house. This house is on the earth, isn’t it? (b) How come the earth is flat here, but before you said it was round?

(0) Don’t know (4, 7)
(1) Sphere and perspective explanation (20, 42)
(2) Round but we live on flat pieces of land (sphere questionable) (28, 21)
(3) Looks flat on the surface but inside is/looks round (sphere questionable) (3, 2)
(4) Child recognizes conflict but cannot/does not explain it. (8, 14)
(5) Round like a pancake, tire (pancake theory) (1, 0)
(6) Looks round from the outside but inside is/looks flat (inside the sphere theory) (24, 20)
(7) Round but we live on flat pieces on top (flattened sphere) (2, 1)
(8) It just looks flat – no mention of size conflict or perspective (Optical illusion theory) (2, 2)
(9) Two earth theory/ they are not the same thing/different places etc. (2, 2)
(10) Child reiterates the Earth is round, does not recognize conflict (also because earth rotates) (5, 0)
(11) Gravity (3, 2)
(12) Child insists Earth is round but cannot explain or gives wrong explanation (2, 1)
(13) Change from round to flat (flat unspecified)
(14) Picture is 2 dimensional/picture makes it look fat. (4, 1)
Other (0, 3)
Missing (3, 0)
So we don’t fall off (1, 1)
Uninterpretable (1, 1)

Model Matching

Sphere – 1, 4 acceptable deviations 2, 3
Hollow – 6, 4 acceptable deviations 2, 14, 17 (these deviations are contingent on context)
Flattened – 5, 7 acceptable deviations 2, 3
Dual – 9 acceptable deviation 2
Disc – 5, 4 acceptable deviations 12, 14

Q9 If you walked and walked for many days in a straight line, where would you end up?

0 Don’t know (2, 0)
1 Back where you started from or in a circle (54, 74)
2 You just keep going, or you end up in a specific place (43, 28)
3 Your get to the end/edge (5, 5)
4 You fall off (4, 7)
5 Other (1, 0)
6 Missing (0, 1)

Model Matching

Sphere – 1 acceptable deviation 2
Hollow – 3 acceptable deviations 2, 4
Flattened – 1 acceptable deviation 2
Dual – 2, 3, 4
Disc – 3, 4 acceptable deviation 2

Q10 Does the Earth have an end or edge? Could someone fall off the end or edge of the Earth?

Why? Where would they fall?

0 Don’t know (2, 0)
(1) No end/edge because the Earth is round, or because the earth is moving. You can’t fall off because the Earth is round, or because the earth is moving (also count if child says the earth is not a square)  (55, 67)

(2) No end/edge because you come back to where you started. You can’t fall off because the you come back to where you started.  (4, 4)

(3) No end because if there were you’d fall off.  (5, 10)

(4) No and no explanation  (16, 11)

(5) No and/but the Earth is in the sky

(6) Yes, there is an edge, but you can’t fall off because of gravity  (5, 5)

(7) Yes, there is an edge, but you can’t fall off because your arms and legs will hold you

(8) Yes, there is an edge, but it is on top of you and you have to go up to reach it.  (1, 0)

(9) Yes, there is an edge, but you cannot get there, or there is a barrier.  (6, 2)

(10) Yes, and you would fall to the ground, earth, dirt, water

(11) Yes, and you would fall into space, sky, moon, stars, celestial objects (6, 9)

(12) Yes, and you would fall down  (1, 1)

(13) Yes, and you would fall into monsters, neverland, magic  (1, 0)

(14) Yes, and you would fall into fire, hell, down there

(15) Yes and you would fall on top of houses

(16) Yes with no explanation/don’t know where you fall  (4, 3)

(17) Other  (4, 3)

(18) Missing data  (2, 2)

(20) Uninterruptable

Model Matching Sphere – 1, 2, 3 acceptable deviations 4, 6
Hollow – 1, 3, 8, 9 acceptable deviations 4, 11, 12, 13, 16 (these deviations are contingent on context)

Flattened – 1, 2, 3 acceptable deviations 4, 6

Dual – 5, 10, 11, 12, 13, 14

Disc – 10, 11, 12, 13, 14 acceptable deviation 6, 7, 9

Q11) (a) Now I want you to show me where Bloomington/Normal is on your picture. (b) Now I want you to show me where China is on your picture

(1) Both inside circle (81, 82)  (2) One is on other non-visible side of earth (22, 30)

(3) One or both are outside circle (1, 0)

(4) Champaign is inside the circle but the child does not know where China is (4, 2)

(5) Does not know where to place either Champaign or China

(6) Other (2, 0)

(7) Missing (0, 1)

Model Matching

Sphere – 1, 2 acceptable deviations 4, 5

Hollow – 1, 4

Flattened – 1, 2, 4 acceptable deviation 5

Dual – 1, 3, 4

Disc – 1, 4 acceptable deviation 5

Q12) Now tell me what is down here below the earth.

(0) Don’t know (2, 2)  (1) Sky, atmosphere, clouds (11, 14)

(2) Celestial objects (49, 41)  (3) Space (45, 67)

(4) Dirt, ground, dust, soil, trees, sewer, lava (8, 5)

(5) Monsters, magic, Neverland  (6) Devil, hell, fire
Q13) Does anyone live, here, on the bottom of the Earth? Why don’t they fall off?

(0) Don’t know  (0, 1)

(1) Yes and wouldn’t fall because of gravity (46, 66)

(2) Yes and wouldn’t fall because the Earth is round, circle or moves  (16, 15)

(3) Yes and wouldn’t fall because the Earth is flat  (5, 4)

(4) Yes and wouldn’t fall because we live inside the Earth or in the ground (16, 15)

(5) Yes, and wouldn’t fall because they are inside houses  (1, 0)

(6) Yes and wouldn’t fall – confused explanation  (6, 5)

(7) Yes and wouldn’t fall – no explanation  (3, 0)

(8) No, but if they did they wouldn’t fall off (gravity)

(9) No, because if they did they would fall off  (3, 2)

(10) No, because it is slippery/pointy/on the bottom (4, 0)

(11) No, because it is/ they are upside down.  (11, 0)

(12) Other  (4, 3)

(13) Missing data  (6, 3)

(20) Uninterruptable
Model Matching

Sphere – 1, 6, 7 acceptable deviations – 2, 8

Hollow – 4, 9, 11 acceptable deviations – 1, 2, 3

Flattened – 1, 6, 7 acceptable deviations – 2, 8

Dual – 2, 3, 6, 7, 8, 9, 10, 11

Disc – 8, 10, 11 acceptable deviation – 1, 2, 3

Q14 – does not get separate scoring. Count anything they say here that is relevant to an earlier question as being part of that answer but requiring focused prompting.

Do not code as part of model building. However, if the student gives an answer to another question in the interview (i.e. 11) than add that information to the student’s answer to that question.
Three children, Dantrell, Teresa, and Eve, wanted to build a snowman. They walked to the end of the street in their heavy coats, caps, and gloves. Even though it was cold, the children were soon warm after running after each other and jumping in the piles of snow that gathered at the edge of the hill at the end of the street.

The hillside was dotted with small leafless trees that dropped little showers of snowflakes as the children ran past. The children started by packing snow into tight snowballs, and then slowly rolled them down the hill. Soon they had three big snowballs that they placed one on top of the other. They packed snow around the bottom one to make it as large as they could. They put two twigs in the middle snowball for arms. They put two big blue paper circles on the top one for eyes and three paper snowflakes to look like buttons in the middle. Teresa brought a little hat and scarf that didn’t fit her anymore. She put these on the snowman. Finally, Eve finished the snowman by giving him a carrot nose, which made everyone laugh. All three children were very proud of their snowman.

Dantrell suggested, “Let’s get one of my father’s old coats to put on the snowman.”

Eve thought that was silly and giggled, “Why does a snowman need a coat? He doesn’t need to stay warm; he’s a SNOWMAN. He’s supposed to be cold.”
“Well, the coat will keep the snowman cold. When it gets warm outside the coat will keep our snowman from melting” Dantrell answered.

Teresa was not sure what to think, “I don’t think a coat will make the snowman cold or hot. I think coats only work for people, don’t they?”

Eve was quick to answer, “Coats are warm. Why else would our mom’s make us wear them? They will heat up the snowman and then he’ll melt right away.”

“I don’t think so,” said Dantrell. “I think that coats keep heat from going through them. They are like a wall for heat. People are already hot. In winter coats don’t let the heat from people leave into the air, so they stay warm. Snowmen are cold. Why can’t a coat keep the heat from the air out, so that the snowman will stay cold?”

“Well, why do people say that a coat is warm, or a sweater is warm?” Eve challenged.

“And why don’t people wear coats when it’s hot outside?”

“Because people are already hot. They make their own heat. They’re warm-blooded. They don’t want to keep the hot in if it’s a hot day.”

“I’m so confused,” complained Teresa. “My coat always feels warm when I put it on.”

“Yeah, but it’s like when we tried to keep that snowball last spring. Remember how your mother put the snowball in a cooler. It was a big thick box that covered it up. It is just like a coat; it keeps the heat out.” Dantrell reasoned.
“That’s not the same thing,” Eve said. “That was a cooler, not a coat. It’s different. That’s why they call it a cooler.”

What do you think? Should the children put a coat on the snowman?
“Summer time is the best,” John said grinning.

“It sure is,” Sam agreed.

John, Sam, and Caleb had wanted to start fishing in the early morning when the fish were hungry and easier to catch. They had gotten to the lake very early, but they had to fill their small metal boat with their rods, bait, and the snacks they had gotten from the gas station. Then they had to drag the boat from the shed out to the water. The boat was not too heavy when it was empty, but with all of their supplies it took some work to pull it into the water.

Although they did not start as early as they had wanted, it was a fine day for fishing. The sky was pale blue and scattered with wispy clouds. The water in the lake was calm. Once John, Sam, and Caleb had floated past the rocks, they could see the reflections of the green trees all along the shore. Each of the boys happily put a small brown worm on the end of their hooks. They cast their lines not far from the shore where tall weeds grew up from the soft mud. This was often a great place to catch fish that liked to hide there and eat small plants and animals. The boys
quickly caught a few small perch, but for awhile they did not catch anything. They had talked and floated, and watched the dragonflies, but now they were ready to catch something bigger. They wanted to bring back a big fish so they could show off to Caleb’s bigger brother Ben who was the best fisherman they knew. What they really wanted was a big catfish.

When they had drifted past the shade of a big tree, Sam looked out toward the middle of the lake. After thinking about the great taste of spicy fried catfish, Sam said, “It’s getting hot; maybe the fish are moving out to deeper water where it’s cold. Let’s go further out and see if things are better.”

“Good idea,” said Caleb. “We’ll put a weight on the line and see what we can catch along the bottom.”

John was silent for a little while. “I don’t know if that’s such a great idea. We have so much stuff in the boat, and there are three of us. The boat is already really heavy. If we go out further we could sink. We’d lose all of our stuff and then we’d have to swim back!”

“Why would we sink if we went out further?” Sam asked.

“Well, the deeper the water, the more the boat has to float. If we go out too far, the boat will sink lower and lower in the water. If we go out far enough, the water is going to go over the top of the boat,” John said in a concerned voice.

“What are you talking about?” Sam asked in frustration. “The water keeps the boat up. The more water there is, the more the boat is going to float. You have nothing to worry about.”
“I don’t know,” said Caleb. He paused for a moment. “Actually,” Caleb continued, “I don’t think it matters one way or another. The boat floats the same no matter how much water there is. If it floats where it’s shallow, it will float the same where it is deep.”

“If that’s true why do they always say that it’s more dangerous to swim in the deep end of the pool?” asked John.

“That’s only when you don’t know how to float. Think about a bathtub. You don’t float in a bathtub, do you? It’s not deep enough. You get in, and you push the water out of the way. The water goes up the sides of the tub, and you go to the bottom. You need deeper water before you can float, because you need more water to lift you up.” Sam explained.

“Yes,” Caleb said, “but I’ve swam across a pool. It had a shallow end and a deep end. I floated just the same all the way across. There just has to be enough water to float and then it doesn’t really matter after that.”

“I just don’t know. And I don’t want to find out when we lose all of our stuff.” John complained.

What do you think? Is John right, will the boat sink if it is in the deeper water?
Appendix H

Transcript of a collaborative reasoning discussion about the story A Coat for Mr. Snowman

Speaker: Utterance

Teacher: O.K. here’s the thing. Some things we’re going to focus on today after we get our discussion started. We are going to focus on giving each person an opportunity to talk. That mean’s watching their non-verbal cues. If they look like they want to say something. [Teacher demonstrates by raising her hand] This would be a non-verbal cue [laughs]. Although we are not required to raise hands, but if somebody does that’s a big hint. O.K. But even little things like this [demonstrates sitting forward and opening her mouth as if to speak] and then sitting back in your chair are cues. Has somebody done that? That might mean they have something to say but somebody else started talking. Let’s honor those cues. Secondly, I really want you to listen to what other people are saying, because with today’s assignment you are going to have to write and answer to the question and that means you are going to use our R.A.C.E. Restating the author’s questions, using the author’s words. Connecting the things you know from outside. And giving as many examples as you can extending it. Ok, so as you’re listening you might have your whole paper just sitting here listening to this discussion. [stops to reprimand two students who are having side conversation] The other thing and this brings me to the fact that I am not going to continue refocusing the group. You guys are in charge of that. If we don’t get through all the points it gives you less information to use in the writing. So you guys need to keep refocusing each other. [Teacher has conversation with one child about forgetting his paper] Ok, right now we’ll go ahead and get started. What do you think, should the children put a coat on the snowman?
Barbara: I don’t think that they should because coats are like, they like make things warmer.

Amil: Yeah they should put it in a cooler.

Gail: A cooler?

Barbara: Who would put a snowman in a cooler?

Gail: Unless you had a really big cooler.

Barbara: Or a very small snowman.

Sam: Or you could just put it in a freezer

Barbara: Freezers are too small.

Amil: OK, so getting back. I think they shouldn’t put the coat on the snowman because coats keep us warm and if the snowman gets warm it basically just melt. So if like if they keep it in a cooler. Like it’s cold in a cooler, right? So um,

Gail: That’s why there’s the word cool in cooler.

Amil: OK, so if they put it in a cooler, and like it won’t let any heat come in so the snowman will want to not melt. So.

Teacher: So let me clarify, are you saying. Because the question is should the children put the coat on the snowman. So your answer, what is your answer?

Amil: No because the coat -

Barbara: They heat up.

Amil: Yeah like because they have cotton or something in them that -

Barbara: That heats up quickly.

Amil: Yeah.
Gail: And what I put is you shouldn’t put a coat on a snowman because the coat is made to block out the cold. And they want the snowman to stay cold so it won’t melt. It just doesn’t make any sense unless you want it melt.

Sam: It already is cold. It already is cold. If they put the coat. I want – My answer is yes they should.

Barbara: Even if it is cold the sun will heat up the coat.

Sam: It blocks. The whole point of the coat is to block the heat out.

Barbara: Yeah but if you wear a coat in the summer it’s not going to keep the cold in.

Sam: That’s because you aren’t cold. We’re warm blooded. We adjust to the climate.

Amil: So if like if I was to open the freezer and come out an hour later, and I’m like freezing cold and I wear a coat all that umm -

Barbara: Yeah, you’re going to get warm.

Sam: Yeah because once you’re out the temperature is warmer and plus you would most likely get frost bite.

Gail: And you’d have to cut off your arm.

Barbara: Off topic.

Sam: I disagree with you [referring to Amil] because if you put a snowman in a cooler most likely it wouldn’t turn to ice.

Amil: Yeah but that will.

Gail: An ice snowman. There is snow in snowman, snow. Ice man?

Teacher: So what is the question that right now you guys are debating, there’s a lot of -

Barbara: You should not put a coat on a snowman. It’s wrong. Plus the sticks would fall off. Like first the sticks -
Sam: But my grandpa when he makes snowman, he just packs snow all the way around the sticks. So they are like snow arms [laughs].

Teacher: Let me ask you this. You guys have given a lot of opinions that you have obviously brought with you. Are there any things in the story that would help you to go for or against what is being said here. Amil, did you hear what I just asked?

Barbara: Right here it says, “That’s not the same thing,” Eve said. “That was a cooler, not a coat. It’s different. That’s why they call it a cooler.”

Teacher: Which page were you on Barbara?

Barbara: I was on page three.

Teacher: Page 3. So let’s look at what Barbara is talking about. So in order to decide if you agree with Barbara or not, what would we have to know?

Amil: Wait, I think that they shouldn’t put a coat on the snowman because what I’ve done in the past when I was a little kid I used to like shoving big balls like snow, I used to put snow on my coat to see what would happen and it just melts in my coat.

Sam: Because you’re body heats up.

Grave: That is true.

Amil: I know, but even though. Because one time I accidently got my coat in the snow. Like a lot of snow came into my coat, and it just melted.

Sam: Was the coat zipped up?

Amil: No.

Sam: If it went into the sleeves because the sleeves are warmer than just the snow.

Amil: It went into the white part.
Gail: Well if it melted because if you wore it and then it still had some heat that you gave onto it and then you put the snow on it, it could still melt.

Amil: I know but my -

Barbara: Plus the snow would soak into the coat.

Amil: The closet that I put my coat in was cold that time. We just got it fixed. And here when we came here it was cold too, so we got that fixed so my coat was in my closet cold, so after I wore it I felt like cold for a long time then I started to warm up.

Sam: On page two in the fourth paragraph, “In winter coats don’t let the heat from people” Wait. [long pause] “Because people are already hot. They make their own heat. They’re warm-blooded. They don’t want to keep the hot in if it’s a hot day.” That’s the fifth paragraph.

Teacher: So how does that support or go against what your answer is.

Sam: It says that people make their own heat but the coat is just trapping it in so they stay warm instead of just letting it out which would happen if they didn’t wear a coat.

Teacher: OK.

Amil: What. Do you put a snowman in a cooler and like freeze and it’ll stay longer. Like as if you put water in like the ice cube tray and put them in the freezer and bring them back out and you leave it out it will like stay longer instead of just when the snow falls and you put out your hand and it just disappears.

Barbara: Desirae, Achira do you have an opinion because I want to know it.

Desirae: I don’t have an opinion.

Teacher: You don’t have an opinion whether a coat would be a good thing to keep on the snowman.
Desirae: I think it’s not, because it would probably melt the snowman.

Teacher: You think it would melt the snowman, why do you think that?

Desirae: Well, because the snowman would get hot and it might melt.

Amil: What’s your opinion Achira?

Achira: I would want, I think you shouldn’t put the coat on the snowman [long pause] because it would melt.

Teacher: Is there anything else in the story that -

Sam: We should test it. Like make one.

Amil: Yeah, we should test it.

Gail: Make baby snowmen.

Sam: Make two snowmen put one coat on one and leave the other without a coat and see which melts first. And make the snowman the same size

Gail: An experiment.

Teacher: That would be a good way to test it.

Barbara: It would be really hard to make it the same size.

Gail: [Inaudible, said to Barbara]

Teacher: So you’d have. So you’d want to make them the same size.

Barbara: Yeah but the same size, I still think it would be hard.

Amil: Wait so ?

Teacher: I see a lot of the points in here that people brought up but I haven’t heard other people address them. For instance, way back on page three Barbara brought up this topic comparing the cooler and the coat but I heard nobody really expand upon that idea. Is a cooler and a coat the same thing? Are they not the same thing?
Barbara: A cooler keeps things cold, a coat keeps things warm.

Amil: I sort of agree with Sam because -

Sam: Or cold because they keep air -

Barbara: No it keeps it warm or makes it warm.

Gail: [Inaudible to Barbara]

Amil: I sort of -

Barbara: A cooler keeps thing warm or makes something. I mean a cooler keeps something cold or makes something cold. A coat makes something warm or keeps something warm.

Teacher: Do you all agree with that?

Sam: I don’t.

Amil: Well, I sort of agree with Sam a little bit now, because if you really think of it, the coat gets wet because of the snow so that and the snow’s cold water so when you put the cold water on the snowman it might have a good chance of -

Sam: It just makes more snow.

Amil: What?

Sam: It has more snow because -

Barbara: It’s not all wet, it’s dry and it’s warm because it is sitting in a closet and you put it on the snowman and it makes it warm.

Amil: I know but to like as it makes it warm the water will go into the coat right? So then the coat will get cold and then like there -

Barbara: No it melts on the coat.

Gail: It melts on the coat and makes the coat wet and -

Barbara: Then it might drip out then
Amil: Yeah but there is some water.

Gail: And the coat absorbs.

Sam: Actually it keeps it cold because if it is already cold it just keeps the heat out of it. And it keeps the cold.

Amil: Yeah, it takes a little bit, it takes a little bit of time, a few minutes to cool down because-

Gail: I see.

Teacher: What do you see that he means?

Gail: Well, if you like put a snowman with a coat on it the coat on the snowman, and the snowman doesn’t have any like natural heat inside it so it won’t like make it warm, and the coat warm and just make - depending on the weather, whether if it’s cold or not. If it’s cold and it will just stay, but if you put it on a human, a warm-blooded person, the heat will spread to the coat which will make the coat warm and will in turn make you warm too.

Sam: Desirae do you have anything to say?

Teacher: What do you think about what Gail just said? Do you think that’s true that the reason the coat is warm on you is because of the warmth you have?

Desirae: Wait, what is the question again?

Teacher: What do you think? Gail just explained that she thinks that when you put a coat on.

Gail can you repeat what you said.

Gail: [Load Sigh]

Barbara: Wait, what if you’re cold? You’re freezing cold. You’ve been standing out without a coat and you shouldn’t really be not wearing a coat; you should be wearing a coat.

Sam: Yeah, but body heat is still transferred.
Barbara: Yeah but you’re freezing cold and then you put a coat on. It’s going to make you warmer.

Amil: I know but -

Sam: That’s it keep the body heat that you have.

Barbara: But what if you don’t have any body heat?

Sam: Well, then you don’t have life. You stay cold like the snowman.

Barbara: No, you’ll make, you’ll get warmer.

Amil: Oh yeah.

Barbara: That’s the point of putting on a coat.

Teacher: OK here’s how we are going to close this out. I don’t want to hear your opinion on yes or no at this point. Should he wear it? But I want to hear the arguments we’ve given for and against, and then you are going to have to tell me your opinion when you do the writing. So what are the arguments we’ve heard for whether you agree with them or not. What are the arguments we’ve heard for putting the coat on the snowman.

Gail: Somebody say yes, some people say no. The people who say yes is that just the rest -

Teacher: Wait, wait, wait. We are not deciding who believes yes. What are the things, what are the reasons you should put a coat on the snowman? What are the reasons we’ve heard?

Sam: Because it traps the body heat or coolness where the snowman is.

Teacher: So you’re saying, let me clarify. Are you saying the coat will keep in cold if that’s what it is or keep in heat?

Sam: It will keep in heat if you have, if you are warm. It keeps you coolest if you are cold or you’re not alive.

Barbara: So you are saying it keeps in the temperature.
Teacher: So that’s an argument for putting the coat on. Did you hear any other arguments for putting the coat on the snowman?

Gail: And some people say that if you put a coat on it, a coat on an object or person or any kind of thing, it just makes it warm which -

Teacher: Wait wait, is it an argument for putting it on?

Gail: No

Teacher: Right now we are focusing on the arguments for putting the coat on. So we said that the coat would keep cold in is what one of the arguments was. Are there any other arguments for putting the coat on the snowman?

Gail: Like you said earlier, if you put a coat on something then they would just be warm.

Teacher: So that’s an argument for putting the coat on the snowman.

Gail: That’s what I think what Barbara said.

Teacher: OK.

Amil: Every student has to say.

Teacher: Now tell me the arguments against putting the coat on the snowman. If your answer is no, you should not put it on, what are some arguments to support that?

Barbara: The warmth from the sun is going to heat up the coat and then the coat is going to melt the snowman.

Teacher: OK, so the warmth of the sun will heat up the coat and then the coat will heat up the snowman. Am I right? Is that what you are saying? Are there any other arguments against putting a coat on a snowman?

Amil: Well, a lot people in this group were saying that if you put a coat on a snowman it will melt, because the coat is warm. The heat will cause it to melt.
Teacher: OK, now here is my question for you guys. In discussing this, now you guys should be able to go back, write a good solid paragraph on saying your answer, what you believe using things out of the story and supporting that with your own knowledge. Do you feel you can support your own answers whatever it is whether it is yes or no from this discussion? Did it help you or not help you come to a conclusion?

Sam: It helped.

Teacher: It helped. Desirae what do you think, did the discussion help you will you be able to do a better a job of your writing?

Desirae: Probably.

Teacher: You think so, some ideas you can use. Barbara what do you think?

Barbara: [Shrugs]

Teacher: Did it help you understand that when you write something like no, that you have to back that up?

Gail: You have to explain it.

Teacher: You have to back it up. So that’s what you have to do in your writing. You can’t just give an answer and just expect whoever’s reading it prove it. You have to prove it. OK? Alright you guys may go back.
Appendix I

Target Story

What Shape is the Earth?

Jill and Maria are talking after school. They are waiting in the gym until the bus arrives. Their teacher, Mrs. Olson, taught a lesson earlier that day about the shape of the Earth. The teacher said that the Earth is round. Jill knows that the teacher thinks that the Earth is round, but she does not believe it.

“I don’t know what the teacher is talking about. It sure looks flat to me,” Jill said to Maria.

“It only looks flat. It’s really round,” Maria said, trying to assure her friend.
“Yeah, but look, the gym is flat. Watch!” Jill took her textbook and placed it carefully on the floor of the gym.

“See, this whole gym is as flat as a book. How can the whole gym be flat if the Earth is round?”

Maria looked a little worried. She wasn’t sure if she understood everything her teacher had said earlier. However, Maria decided she had to try to explain it to Jill anyway. “Jill, the Earth is really really big. Some parts look flat and some parts look bumpy. The gym looks flat and places like the hills in town look bumpy. But that’s because they are only really small pieces of the Earth. You would have to see a super big piece to see that the Earth is round.”

“But even the lake looks flat and that’s pretty big. It’s a lot bigger than a gym.”

Maria felt stumped, but then she saw a big pumpkin left over from last week’s Halloween party. “Look at this, Jill. See this small piece of the pumpkin. If I bring it up
close to my face, it looks flat because it is only a small piece, but if I stand back, I can see it is really part of a round pumpkin.” Maria held out the small piece of pumpkin smiling.

Maria’s smile faded when she saw how confused Jill looked. “I don’t know,” Jill said. “Even from the top of the tallest building in the city, it still doesn’t look round.”

Maria remembered her teacher talking about this. Maria took out her textbook. She flipped through the pages to the part they had used in class. She held up the book to Jill, and showed her a picture of the Earth taken from the space shuttle.
“See Jill, once you get far enough from Earth you can see that it is round. But you need to go really far, like in a rocket. When you get really really far you can see that it is round like a ball and that we live on the outside.”

“That’s another thing I don’t understand. How can anyone live on the outside of a ball? Everyone on the bottom would just fall off! If it is round, we could only live on the tiny part at the top.” Jill reasoned. The two girls went into Mrs. Smith’s class and borrowed her globe and some toys to experiment. “Alright Maria, watch what happens to this car if I let go of it.” Jill held a yellow car near the bottom of the Earth. Just as she said it would, the car fell to the floor as soon as she let go.
“O.K. I’ll admit that was a difficult part of the lesson today. But Mrs. Olson said that gravity doesn’t really pull everything down. It pulls everything toward the center of the Earth. So no matter where you are, you get pulled toward the ground. You don’t fall off.”

“Yeah, I remember that part, but that’s just weird. Everything falls down when you let go,” Jill said.

“Well, imagine that gravity is sort of like a magnet?” said Maria.

“What does that have to do with it? Gravity makes things fall. It’s not a magnet.” Jill argued.

“Yes, but gravity is like magnets. It pulls everything in. Remember the magnet Earth Mrs. Olson showed us?” Maria asked. Maria put the toy cars on the magnet Earth and they didn’t fall off. Even the one near the bottom did not move.
“See,” said Maria. “The Earth’s gravity is sort of like a magnet. It pulls everything toward the center no matter where you are.”

Just then two buses pulled up. Maria and Jill had to take different buses. Jill yelled, “See you tomorrow, Maria. I’ll think about it tonight.”

“Me too! See you, Jill,” Maria yelled back.