Learning and Instruction 33 (2014) 67–80

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

The effects of reading to prepare for argumentative discussion on cognitive engagement and conceptual growth

Brian W. Miller[a,*,1], Richard C. Anderson[a], Joshua Morris[a], Tzu-Jung Lin[a,2], May Jadallah[b], Jingjing Sun[a]

[a] Center for the Study of Reading, University of Illinois at Urbana-Champaign, 51 Gerty Drive, Champaign, IL 61820, USA
[b] Illinois State University, Box 5330, Normal, IL 61790, USA

ARTICLE INFO

Article history:
Received 6 June 2013
Received in revised form
11 April 2014
Accepted 13 April 2014
Available online 8 May 2014

Keywords:
Conceptual change
Science learning
Argumentative discussion
Engagement

ABSTRACT

Dialogue based approaches to education have been shown to benefit students through the quality of shared discourse. Warm conceptual change theories propose that these benefits are also mediated by increasing student engagement. Discourse and engagement effects were isolated in this study by having 130 third and fourth grade students read a science text for different purposes (no stated purpose, to prepare for a regular classroom discussion, or to prepare for an argumentative discussion) and then testing children before the discussion took place. Children who anticipated a discussion, especially an argumentative discussion, read more slowly than other children after controlling for fluency. A subset of reading times predicted conceptual growth. Finally some children who participated in argumentative discussions had higher rates of conceptual growth. Results substantiate the efficacy of argumentative discussion as a context for reading scientific texts, and they support the central mechanism of dual-processing theories of warm conceptual change.

1. Introduction

Research over three decades has established social interaction as a powerful pedagogic tool in the science classroom (Bennett, Hogarth, Lubben, Campbell, & Robinson, 2010; Hofstein & Lunetta, 2004; Lemke, 1990; Solomon, 1993). In particular, argumentative discussion has been shown to be an effective method (Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Kuhn, 1993; Osborne, 2010; Pontecorvo, 1987; Zohar & Nemet, 2002). Jimenez-Aleixandre and Erduran (2008) gave the following list as a summary of the ways that argumentative discourse improves science education outcomes: helps students use the cognitive processes of scientists, develops critical thinking, teaches the discourse of science, communicates the culture of science, improves reasoning about scientific topics, and leads to better understanding of the nature of science. All of these effects of argumentative discussion are valuable, but the list focuses solely on the cognitive benefits of exchanging words in an argumentative discourse and not on the affective and motivational impacts of these social interactions. Anyone invigorated by a good discussion can sense that social approaches can change the way people feel and think above and beyond the words that are exchanged. The affective and motivational aspects of discussion are rarely studied despite their potentially important practical implications. In a review of 94 studies of small group discussion in science classrooms, Bennett et al., (2010) found no research focused on affective or motivational consequences.

Of the many attitudinal and motivational constructs that might be enhanced by discussion and reading for discussion, engagement is perhaps the best suited to the study of the effects of discussion, because it is situational rather than trait oriented. Fredricks, Blumenfeld, and Paris (2004) describe engagement as a multifaceted construct consisting of three interconnected aspects: behavioral engagement which includes students actively participating in learning activities, emotional engagement which includes having positive feelings about learning activities, and cognitive engagement which includes "the willingness to exert the effort necessary to comprehend complex ideas and master difficult skills" (p. 60)

* Corresponding author. Towson University, Elementary Education Department, 8000 York Road, Towson, MD 21252-0001, USA. Tel.: +1 410 704-4876; fax: +1 410 704-2063.

E-mail addresses: bwmiller@towson.edu (B.W. Miller), csrrca@illinois.edu (R. C. Anderson), morris41@illinois.edu (J. Morris), tlin27@illinois.edu (T.-J. Lin), mjadall@ilstu.edu (M. Jadallah), sun50@illinois.edu (J. Sun).

[1] Now at Department of Elementary Education, Towson University, USA.
[2] Now at Department of Educational Studies, Ohio State University, USA.

http://dx.doi.org/10.1016/j.learninstruc.2014.04.003
0959-4752/

encountered during activities. All three of these types of engagement interact and are likely affected by school activities.

The study of the connections between, engagement, science learning, and social interactions such as argumentative discourse can be explained in terms of a theoretical perspective known as *warm conceptual change*. The field of conceptual change developed to address the repeated observation that students' non-scientific preconceptions about the natural world are not easily changed (Duit & Treagust, 2003). The original theories of conceptual change foregrounded cognitive processes needed to make sense of scientific principles and relegated motivational, affective and social factors to the role of minor moderating variables (Pintrich, Marx, & Boyle, 1993). The explanatory and pedagogic shortcomings of the original theories led to new theories termed 'warm' because they gave motivational, affective, and social variables a central role (Abd-El-Khalick & Akerson, 2004; Cobern, 1996; Linnenbrink & Pintrich, 2002; Pintrich et al., 1993; Sinatra, 2005).

1.1. *Persuasion based theories of warm conceptual change*

Some of the warm conceptual change theories are directly influenced by theories of persuasion developed in the fields of social psychology and communications (Dole & Sinatra, 1998; Gregoire, 2003; Mason, 2001; Murphy, 2007; Woods & Murphy, 2001). There is a natural resonance between persuasion and warm conceptual change research, because both address the often emotionally-laden phenomenon of people changing their minds – often requiring them to abandon well-entrenched beliefs.

One prominent persuasion based conceptual change theory is the Cognitive Reconstruction of Knowledge Model (CRKM) developed by Dole and Sinatra (1998). The CRKM and other similar theories make use of the dual-process theory of persuasion known as the Elaboration Likelihood Model (ELM, Petty & Cacioppo, 1986) as well as the classic conceptual change model (CCM, Posner, Strike, Hewson, & Gertzog, 1982).

The Elaboration Likelihood Model (Petty & Cacioppo, 1986) theorizes that when a person receives a message, but before much substantive information has been acquired, the person must choose (consciously or without awareness) how deeply to process that message (Craik & Lockhart, 1972). While people generally want to understand the world and to hold accurate beliefs, they are also reluctant to expend cognitive effort unnecessarily. The tension between these two desires determines the depth of processing (similar to other theories of setting standards of coherence in reading; see van den Broek, Lorch, Linderholm, & Gustafson, 2001). The ELM is called a dual-process theory, because the extremes of processing imply qualitatively different cognitive strategies. Relatively shallow "peripheral" processing includes quick cognitive strategies such as counting the number of arguments. Relatively deep "central" processing includes cognitively intensive strategies such as analyzing the logic of arguments. While attitude change is possible through either route, change through the central route will be stronger and more persistent over time because it results from more elaborate strategies (Petty, Rucker, Bizer, & Cacioppo, 2004).

In constructing the CRKM, Dole and Sinatra (1998) tried to preserve the best features of both the ELM and classical theories of conceptual change. The CRKM has a dual-processing structure similar to the ELM. The theory reflects earlier ideas about conceptual change with the addition of key personality and motivational constructs including social context. One important difference between the ELM and the CRKM is that the central mechanism is depth of engagement rather than depth of processing. In particular the CRKM focuses on the cognitive form of engagement. In addition, the structure of the CRKM is explicitly described as an iterative process in which there is a continuous flow of receiving information, engagement and processing. The CRKM does not claim to be exhaustive and is still being developed through empirical study.

Early studies that invoked the CRKM tended to be exploratory instructional intervention studies (Alexander, Fives, Buehl, & Mulhern, 2002; Nussbaum & Sinatra, 2003). More recently researchers have probed specific predictions of the theory. Some of these studies have examined the interplay between several warm variables over large and intermediate instructional time scales. These researchers have either not needed to gather direct measurements of engagement (Broughton, Sinatra, & Nussbaum, 2011; Lombardi, Sinatra, & Nussbaum, 2013; Taasoobshirazi & Sinatra; 2011) or have used off line measures such as self-report (Hynd, 2003; Johnson & Sinatra, 2013). Other researchers have concentrated on the moment-by-moment effects of individual variables on engagement over short time scales. For example, Ranellucci et al. (2012) used think-aloud protocols while students read two short refutational reading passages to explore the interplay of student's goal orientation, depth of processing and conceptual change.

1.2. *The current study*

The present study employs the CRKM as a theoretical framework for exploring the effects of discussion and particularly argumentative discussion on students' cognitive engagement while reading a science text. Researching this aspect of discussion poses a series of difficult methodological hurdles. First and foremost, it is difficult to isolate the effects of the words exchanged during discussion from the effects of engagement. However, it is possible to infer what the effects of cognitive engagement in discussion might be by testing the effects of the mere anticipation of discussion. This would be an important result in itself, and it would suggest that students increased engagement preparing for a discussion might continue throughout the lesson.

A second methodological problem is that engagement is a difficult construct to measure because it is multifaceted and can be observed at different time scales (Sinatra & Heddy, 2013). Since this study focuses on the anticipation of a discussion, which is likely a subtle and short-lived phenomenon, it uses sensitive measurement and frequent sampling. In addition, the study simplifies measurement by focusing on one aspect of engagement – cognitive engagement. While it is only one aspect of engagement, it is the one most closely associated with dual-processing models and their effect on conceptual change.

1.2.1. *Manipulating anticipation of a future discussion*

In this study, the anticipation of a future discussion is hypothesized to impact a student's purpose for reading. Reading to prepare for a discussion is a method of increasing cognitive engagement that dates back to early research in persuasion (Chaiken, 1980). Reading for the purpose of discussion is likely to change the experience of reading in a number of ways including changing the feeling of accountability to peer judgment (Johnson & Eagly, 1989). One problem with instructing students to read for the purpose of preparing for a discussion in the context of classrooms rather than a laboratory context is that classrooms have very different forms of discussion.

Researchers studying classroom discourse have observed that some classrooms have a format for discussion during which the teacher does the majority of talking, controls the content, and has evaluative authority (Cazden, 2001; Lemke, 1990; Mehan, 1979). Anticipating this type of discussion might have small effects, because students know they will rarely speak and are chiefly accountable to the teacher to answer questions of limited scope.

Preparing for this type of discussion requires no more than a basic level of comprehension of the text.

In contrast to traditional forms of teacher-led discussion, a number of different approaches have been developed to foster more engaging and effective discussions (for example 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, McKeown, Sandora, Kucan, & Worthy, 1996; McKeown, Beck, & Worthy, 1993), and Philosophy for Children (Sharp, 1995)). One such approach is Collaborative Reasoning (CR), which features small groups, open participation, and argumentative discourse about a controversial text (Anderson, Chinn, Chang, Waggoner, & Nguyen, 1998; Clark et al., 2003; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). Students in CR discussions have been found to contribute a higher percentage of talk, ask more questions, and challenge their peers more often than students in regular classroom discussions (Chinn, Anderson, & Waggoner, 2001), and students rate CR discussions as more interesting than regular discussions (Wu, Anderson, Nguyen-Jahiel, & Miller, 2013).

Students preparing for a CR discussion will have an expectation that they must speak, have an opinion, and support their opinions with reasons and evidence. CR discussions are not argumentative in the commonly used sense of an adversarial debate. To the contrary, CR discussions have many of the characteristics Michaels, O'Connor, and Resnick (2008) have termed *accountable talk*. Accountable talk helps to foster a classroom community in which "students have the right to speak and the obligation to explicate their reasoning, providing warranted evidence for their claims so that others can understand and critique their arguments" (pg. 284–285). It is this balance of rights and obligations that makes argumentative discussions like the ones incorporated into the CR approach socially accountable. Beyond basic comprehension, students preparing for this type of discussion must consider how they will take advantage of their speaking rights so that they can have the joy of expressing themselves and how they will satisfy their obligations to their peers to contribute to the collective search for understanding.

To further investigate the effects of the purpose of reading, the present study has three different purposes for reading: no stated purpose, reading to prepare for a classroom discussion, and reading to prepare for a CR discussion. In addition a fourth condition was having no stated purpose for discussion, but participating in CR discussions. This was included, because experiencing CR discussion might have an influence on reading even if it was not announced. The CR instructional framework is more than a form of discussion. It includes reading different types of stories, learning about argumentation, and becoming acquainted with new norms of social interaction. Therefore, it is possible that learning the CR format effects reading and learning beyond the effects of social accountability.

1.2.2. Reading time as an online measure of cognitive engagement

Cognitive engagement is mental effort invested in comprehension, thinking and learning (Fredricks et al., 2004; Wang, Willet, & Eclles, 2011). There is a long history of measurement of cognitive engagement using the time readers spend on words and sentences. Just and Carpenter proposed the eye-mind connection in their 1976 paper suggesting that the amount of time that people's eyes rest on different words during reading indicates different amounts of cognitive effort. The eye-mind connection has been the basis of numerous research findings. For example, participants look at pronouns longer if the antecedent is difficult to identify (Ehrlich & Rayner, 1983). Readers take longer to read words that have ambiguous meanings (Duffy, Morris, & Rayner, 1988) and words that require inferences (O'Brien, Shank, Myers, & Rayner, 1988). Readers spend more time on interesting sentences than less interesting sentences (Wade, Schraw, Buston, & Hayes, 1993), relevant or important sentences (Kaakinen & Hyona, 2005) and sentences related to the purpose for reading (Reynolds & Anderson, 1982; Rothkopf & Billington, 1979). Reading time has been assessed during the reading of science texts to evaluate cognitive engagement. Several researchers have found that students who undergo conceptual change have slower overall reading times than other students reading the same texts (van den Broek & Kendeou, 2008; Mikkilä-Erdmann, Penttinen, Anto, & Okinuora, 2008; Penttinen, Anto, & Mikkilä-Erdmann, 2013). All of this research combined supports the contention that the more thinking is required to understand a piece of text, the more slowly people will read it.

Reading times reflect all of the processes involved in reading from perception and decoding through comprehension and finally to reflection and evaluation. To use reading times to measure cognitive engagement, it is necessary to experimentally or statistically control for the time taken by other processes than reflection and evaluation. The construction-integration model of discourse comprehension (van Dijk & Kintsch, 1983; Kintsch, 1998) is a useful theoretical framework for doing so because it organizes the many variables into three levels of discourse representation — surface-level, textbased, and situation model. Cognitive engagement is reflected in the remaining variance in reading times after accounting for the variance associated with other reading processes as reflected in the three levels of discourse representation.

Even after accounting for other sources of variance in reading times, there is the possibility that the extra time is a result of distraction, daydreaming, or other processes. The influence of momentary distractions can be controlled by trimming reading times that are out of range, but reading times need to be combined with other measures to distinguish between cognitive engagement and pervasive distraction (Mikkilä-Erdmann et al., 2008). In this study, a test of basic text comprehension was added for this purpose. If students read slowly because they are distracted, comprehension would suffer. If students read slowly because they are engaged then their comprehension would be better than the distracted reader. Also added was a test of reading fluency to control for stable individual differences among students in the time it takes them to read with comprehension.

1.2.3. Conceptual change and growth

The CRKM explicitly postulates that there is a continuum of conceptual change ranging from strong to weak describing the degree that the knowledge representations must be restructured. Chi (2008) describes three points along this continuum. At one end are single ideas that can be summarized in a single statement. At the other end are changes in belief requiring shifts in ontology. The intermediate type is a shift in mental models. Due to the brief intervention, this study will focus on the intermediate sized conceptual change – in particular, the child's belief about the shape of the Earth. The term "conceptual change" will be used in this study to refer to changes in a child's entire mental model of the shape of the Earth. When the student is only changing a portion of the model consisting of one or a few ideas, it will be referred to as conceptual growth.

1.3. Summary and research questions

Applying the CRKM to the investigation of the effects of reading for the purpose of preparing for discussion on science learning, suggests a mediation model – reading for the purpose of a discussion will increase conceptual change mediated by increased engagement. This is similar to other engagement-mediated models found in previous work (Greene, Miller, Crowson, Duke, & Akey, 2004; Johnson & Sinatra, 2013). However, for methodological reasons discussed in Section 2.8, instead of exploring this as a single hypothesis as other studies have done, the two parts of the model will be investigated as three separate hypotheses:

- Reading for the purpose of preparing for a discussion increases cognitive engagement in reading a science text, and this effect is stronger when students are anticipating an argumentative discussion (Hypothesis 1). To test this, the purpose of reading was manipulated and cognitive engagement was measured by taking reading time measurements. If Hypothesis 1 is correct, then adjusted reading times should be slower for students reading to prepare for a discussion and slowest for those reading to prepare for an argumentative discussion.
- Increased cognitive engagement leads to increased conceptual growth (Hypothesis 2). To test this, reading times from a selected set of clauses was used to predict growth on a corresponding conceptual question. If Hypothesis 2 is correct than the slower students read these clauses, the higher will be the rate of growth on this question.
- Reading for the purpose of preparing for a discussion increases conceptual change (Hypothesis 3). To test this, students were tested on their conceptual understanding of a science topic before and after reading a science text. If Hypothesis 3 is correct than reading for the purpose of preparing for discussion should increase the rate of conceptual change and should be highest for students reading to prepare for CR discussion.

Together these three hypotheses will support the mediation model. Hypothesis 1 will support the first step of the model using all of the reading time measures. Hypothesis 2 will support the second path in the mediation model, but it relies on only a section of the data. Hypothesis 3 supports the contention that the results predicted by Hypothesis 2 holds for the entire data set.

Answering these research questions could further expand a quickly growing body of research supporting the central contention of the warm conceptual change movement that motivational and emotional states have too powerful an effect on learning to be relegated to the error term in science education research (Mason, Gava, & Boldrin, 2008; Taasoobshirazi & Sinatra; 2011). This study will also further test the usefulness of the CRKM's dual-processing mechanism by inducing engagement in a way different from previous studies and in a younger age group than in any previous study. Finally, it could have practical implications for the use of argumentative discussion approaches such as CR. If argumentative discussion increases conceptual change through increased engagement, that would suggest its use as a context for many types of learning activities such as reading.

2. Method

2.1. Participants

There were 130 students who volunteered for the study with signed parental consent: 31 3rd grade students, average age 9 years 2 months, 12 boys and 19 girls; and 99 4th grade students, average age 10 years 0 months, 58 boys and 72 girls. Participates were in six 4th grade and two 3rd grade classrooms in four different schools in a small Midwestern city. One classroom from pairs of classrooms matched on grade level and school demographics was randomly assigned to the CR condition while the other served as the control.

2.2. Teacher training

The eight participating teachers were two men and six women who had at least three years teaching experience. Teachers attended a three-hour workshop on Collaborative Reasoning that covered how to introduce a CR discussion and frame a single question for discussion. Teachers were told that their primary task was to facilitate discussion by encouraging students to control turn-taking, back their positions with reasons and evidence, and to politely challenge others.

Teachers in the control condition were asked not to begin CR discussions until after the experiment. A discussion in each of the control classrooms was videotaped to document that the discussions were consistent with the usual norms of teacher-led classroom discussion. All of the CR discussions were videotaped. CR teachers were given targeted feedback and advice as needed.

2.3. Pretest measures

The study took place over six weeks. Pretests were given the first two weeks, discussions took place in the next two weeks, and the target text reading and post testing was completed in the final two weeks. Below I report reliability and validity estimates from previous published research when available. Reliability estimates derived from data from this study are reported in the Results section (3.1).

2.3.1. Pencil and paper pretests

There were three group administered pretest measures. In order to control for the variation in student's general reading comprehension the Stanford Achievement Test Series 10th Edition (SAT-10) reading comprehension subtest was obtained from the district. The SAT-10 reading comprehension subtest has a reliability of .88 from a nationally representative sample (Pearson Inc., 2003). Six of the students did not have SAT-10 scores, and their scores were imputed based on the other three pre-tests, age, and gender.

Since the shape of the Earth is a topic highly dependent on students' ability to manipulate a three-dimensional model from different perspectives, a spatial reasoning test was administered to help reduce the variance caused by differences in this ability. The CogAT (Riverside Publishing, 2001) figure analysis subtest was chosen because it asks students to mentally fold pieces of paper, which is an analogous task to imagining the Earth from different perspectives (Shepard & Feng, 1972). All of the CogAT subtests have reliability that exceeds .90 (Spies & Plake, 2005), and the test as a whole is strongly correlated with other tests of cognitive ability including the WISC and the Woodcock-Johnson (Lohman, 2003a, 2003b).

Since reading time measures are partially dependent on students general reading fluency a reading fluency test was included to reduce variance in the analysis of reading times. The reading fluency test (He, 2007; Wu et al., 2013) consists of 110 sentences. Students are instructed to read each sentence and mark it true or false. The reading fluency test consists of a series of sentences students read and mark true or false. The sentences contain easy vocabulary and are designed to be obviously true or false so that students have a very high level of accuracy. The number of syllables in the sentences correctly marked in five minutes constitutes the reading fluency score. No reliability estimates for this measure are currently available.

2.3.2. The shape-of-the-Earth interview

Approximately one week after the group administered pretests, but still prior to the CR discussions, the students were given the shape-of-the-Earth interview as developed by Vosniadou and Brewer (1992). The interview contains questions that elicit both verbal responses and drawings. It includes factual questions such as "What is the shape of the Earth?" or "Show me where China is on your drawing," as well as generative questions that pose novel situations such as "if you walked and walked for many days in a straight line, where would you end up?" These questions were designed to be analyzed as a complete set allowing researchers to identify their overall mental model. The only changes made to the interview were to add two questions: "Does anyone live, here, on the bottom of the Earth?" and if so, "Why don't they fall off?" and if not, "Why not?" These questions were added to allow for a targeted analysis connecting a specific section of the text with a specific question in the interview. Furthermore it addresses a key aspect of understanding the shape-of-the-Earth that was not included in the Vosniadou and Brewer (1992), which is the idea that gravity pulls objects toward the middle of the Earth allowing people to more comfortably inhabit a spherical Earth.

Vosniadou and Brewer (1992) found that most children have mental models of the Earth that are consistent with one of six types shown in Fig. 1: The disc-shaped Earth, the rectangular Earth, the hollow sphere (similar to a fish bowl half filled with dirt in which people live on the inside), the dual Earth (one small uninhabited spherical Earth in the sky and another flat Earth on which people live), the flattened Earth, and the spherical Earth. In addition to these categories, some children have mixed models that contain aspects of more than one model (see Brewer, 2008; Lelliott & Rollnick, 2009, for reviews of research). Brewer (2008) categorized the different models according to their conceptual distance from the spherical model. Based on results from ten cross-sectional studies, Brewer ordered the models according to the peak age at which children held each model: rectangular, disc, dual Earth, flattened Earth, hollow Earth, and spherical Earth. Brewer (2008) did not classify mixed models according to a developmental trajectory.

2.3.2.1. Interview procedures. The interviewers received training in techniques that might clarify students' answers while being careful not to ask leading questions. In contrast to Vosniadou and Brewer (1992), interviewers did not receive instruction about children's models of the Earth. The interviews were conducted individually, lasted between 8 and 18 min, and were recorded on digital recorders. Interviewers took note of gestures and other nonverbal communication.

2.3.2.2. Interview scoring. Answers given by a student were categorized according to the scoring rubric developed by Vosniadou and Brewer (1992) with minor changes made to reflect the cultural background of this group of students. Each answer was considered either consistent with a model, an acceptable deviation from the model, or an unacceptable deviation from the model. Vosniadou 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" (p. 554).

Vosniadou and Brewer allowed no unacceptable deviations and at most one acceptable deviation for an interview to be scored as consistent with a model. This criterion was required to identify the existence of specific mental models of the Earth, but the current study is focused on the process of model change. For this reason, we increased sensitivity by creating two categories for each model: strict and loose. The strict designation was given to interviews meeting the original criterion. The loose designation was given when there were no unacceptable deviations but two or more acceptable deviations. For the loose category, claims of ignorance (e.g. "I don't know") and missing information were considered acceptable deviations, but for strict categorization they were not. Interviews that did not qualify as loose or strict instances of any of the expected mental models were classified as mixed.

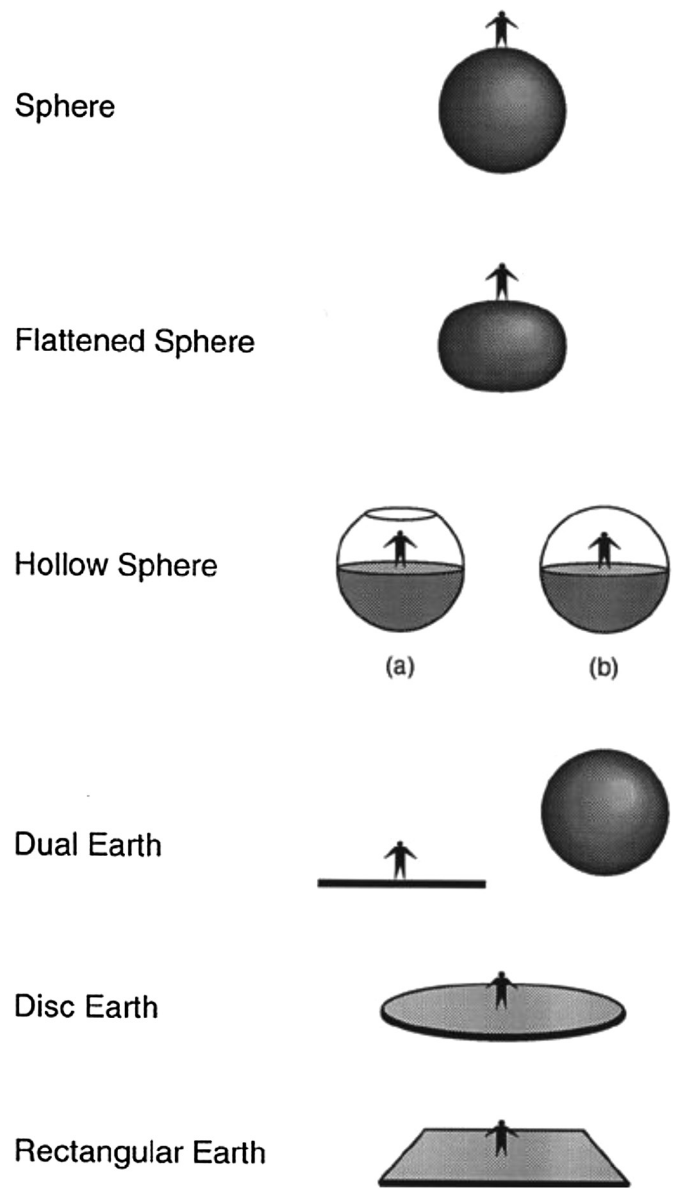

Fig. 1. Six models of the Earth. Reprinted from "Mental models of the earth: A study of conceptual change in childhood" by Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, p. 53. Copyright Elsevier, used with permission by the publisher.

The coding of the interviews was conducted blind to whether the interview was administered before or after the treatments (although a small number of interviews revealed that they were posttests by referring to the story). Conceptual change was defined using Brewer's (2008) developmental order of models. Any student moving to a more sophisticated model or from a loose to a strict version of a model was classified as showing conceptual progress. A 25% sample of the interviews were blind coded by an independent coder resulting in satisfactory interrater reliability (Cohen's kappa = .82).

One challenge to implementing this scheme was that Brewer's order does not include mixed models. These models are diverse in nature and therefore difficult to place within the continuum. Since the spherical model is the scientifically accepted model, any deviations would be considered less sophisticated, and therefore any student beginning with a mixed model and changing to a spherical model was scored as achieving conceptual progress. However, there were seven individuals in the data set that started with a mixed model and transitioned to a hollow Earth model. A reading of these interviews by the researchers suggested that these students progressed from an unsure and confused model with some elements of the hollow Earth model to an articulate hollow Earth model. However, the sample size was too small to perform a reliability estimate. Therefore the analysis addressing Hypothesis 2 (Section 2.8) was run twice: once with the assumption that mixed models are more sophisticated than hollow Earth models and once assuming they are less sophisticated. The substantive results were the same; therefore, the results assuming that mixed models are less sophisticated than hollow Earth models was reported.

2.4. The discussion treatments

The discussion treatment was intended to manipulate students' expectations about what a discussion entails. The four classrooms in the regular instruction condition were asked to continue their usual language arts curriculum including their usual discussions. One discussion judged as typical by each teacher was videotaped. This videotape was used to determine the characteristics of a discussion in the regular classrooms.

The four classrooms in the CR discussion treatment conducted four CR discussions over two weeks. Four discussions were considered the minimum to create a change in expectations based on previous research (Clark et al., 2003). The classrooms discussed the stories in the same order. The first two stories asked ethical and child-friendly policy questions and were chosen to enable a successful introduction to CR. The third and fourth stories were written for this experiment to introduce children to argumentative discussion about a science topic, and to make transfer of expectations to the target story more likely. They were based on concept cartoons described by Keogh and Naylor (1999). They were written in the same format as the shape-of-the-Earth story but with topics (heat transfer and buoyancy) different enough so that specific transfer was not likely. Both science stories were pilot-tested to insure that students in this age range would find the issues controversial and engaging.

2.5. The shape-of-the-Earth text

The target text was written in an argumentative rather than refutational style; it presents both positions rather than directly arguing for a particular position (Miller, 2009). The story tells about two girls presenting reasons and evidence for and against a spherical model of the Earth. Each girl gives two arguments about why the Earth looks flat, and one argument addressing the issue of whether people could live on the bottom of a round Earth (Hayes, Goodhew, Heit, & Gillan, 2003). This type of text allows students with any initial shape-of-the-Earth model to experience arguments in accord with and contrary to their own. The passage consists of 569 words, 112 clauses, and 6 pictures. The passage has a Flesch–Kincaid Grade level of 4.3.

The shape of the Earth topic was chosen because it is appropriate to the age range and is likely to avoid both ceiling and floor effects (Hayes et al., 2003; Sneider & Pulos; 1983; Vosniadou & Brewer, 1992). Chi (2008) reviewed several previous studies and estimated that for topics of this complexity approximately 60% of students with a flawed mental model can attain a correct model from "relatively brief instruction from text" (p. 69).

2.6. Presenting the text and posttest measures

A week after the final CR discussion, all students read the shape-of-the-Earth story and took the immediate post-tests individually using SuperLab Version 4 on Apple Macintosh notebook computers. This program automatically records the time of keystrokes. Students began by reading a practice story, *The Tortoise and the Hare* (adapted from Jacobs, 1902), to familiarize themselves with clause-by-clause reading. A researcher was present to supervise and answer questions.

The shape-of-the-Earth story was divided into clauses. A clause was defined as a distinct part or member of a sentence containing a subject and a predicate ("Oxford English Dictionary online", 2013). A small number of clauses were combined to avoid confusion. The clauses appeared one by one on a single line centered on the screen. When the space bar was pressed, the previous clause was replaced with the new clause. Students could not go backwards in the text. Interspersed within the text were six pictures.

2.7. Posttest measures

After reading the story, 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 text. This test used the Sentence Verification Test (SVT) format (Royer, Carlo, Dufresne, & Mestre, 1996) which is a true and false test designed to discriminate between participants who might have a memory of the surface features of the text but have not retained the gist meaning. The SVT for this story contained 16 items (one question was omitted from scoring because of an ambiguity in meaning). Students were asked to mark "TRUE" if the sentence had the same meaning as something a character had said in the story, or to mark "FALSE" if it had a different meaning from what a character had said.

2.7.1. Post-test shape-of-the-Earth interview

The posttest interview was identical to the pre-test interview, but was administered by a different interviewer. To test for persistence of conceptual growth, half of the students were given the interview immediately after the other measures and half were given the interview after a one week delay. The difference in outcomes between the two interview times was not significant ($t = 1.09$, $p = .86$); therefore, the immediate and delayed interviews have been combined in this analysis. Students not receiving the shape-of-the-Earth interview immediately received the epistemological interview developed by Mansfield and Clinchy (2002). Thus, all students within each class received an assessment each week, reducing student confusion or anticipation.

2.8. Analysis

The CRKM predicts a model in which reading for the purpose of preparing for a discussion increases conceptual change mediated by increased engagement. A mediation model would ideally be tested using a path analysis. However, reading times do not easily enter the Structural Equation Modeling framework. As described in Section 1.2.2, reading times have a naturally high variance especially in children. They require multilevel modeling to accurately model this variance and isolate the effects of specific treatments (Richter, 2006). Entering reading times into a path analysis would require either collapsing the reading times into one number masking any effects with unexplained variance or entering the reading times for each clause in the model exhausting the degrees of freedom. For this reason the mediated model was broken into three hypotheses allowing a statistical analysis appropriate to each hypothesis to be used.

To analyze Hypothesis 1, a mixed-effects model with crossed random effects for students and clauses (Baayen, Davidson, & Bates, 2008) was used. This technique allowed student level variables such as reading fluency and clause level variables such as clause length to be modeled simultaneously creating more power and better model fit (Richter, 2006). The validity of this result was bolstered by the results of the reading comprehension test, which was analyzed using multiple regression.

Hypothesis 2 tested if the reading times for a selected set of clauses would predict conceptual growth on a corresponding interview question. Collapsing the reading times of a small homogenous subset of clauses solves this problem of unaccounted for variance by maximizing the similarity between clauses. The section of the story discussing why gravity allows people to live on the bottom of a spherical Earth is an ideal subset to analyze, because it directly matches the last question on the interview. To further insure that the clauses are homogenous, two analyses were conducted: one for clauses spoken by the character arguing for a spherical Earth, and one for clauses spoken by the character arguing against a spherical Earth. Both of these sets of clauses were used to predict conceptual growth. The dichotomous variable of conceptual growth was given a one if a student had a pretest answer that was inconsistent with a spherical model of the Earth and a posttest answer that was consistent with a spherical Earth. Since the dependent variable was dichotomous these tests were conducted using logistic regression (Agresti, 1996).

Given that the test for Hypothesis 2 used only a subset of the clauses and interview questions a third analysis was used to bolster the contention that the overall model accurately described the complete reading and interview data (Hypothesis 3). To analyze if the reading purpose treatments increased conceptual change, the interview results were transformed into a dichotomous variable of either progress or no progress. These data were then also modeled using logistic regression (Agresti, 1996).

In all cases, covariates played a crucial role in isolating the effects of the manipulations. An effort was made to control for as many sources of variation as possible while avoiding multicollinearity. In order to accomplish this, the final models were selected using a blockwise selection process (Pedhazur, 1997). In this procedure covariates are grouped into blocks with each block representing a theoretically important category. As each block is entered into the model, the non-significant variables are removed and the next block is added (alpha level was set at $p < .05$ for all analyses). For highly correlated variables, only the most theoretically interpretable variables are retained. The final model adds the treatment variables and theoretically important interactions. This process insures that the final model will consist of at least some representatives from each category while still allowing for interpretable results.

3. Results

3.1. Pretest measures

Descriptive statistics for all measured variables (excluding the interview results) are presented in Table 1. The pre-test means were not significantly different between treatment groups except for the SAT-10 reading comprehension scores which were significantly lower for regular instruction classrooms than CR classrooms ($t(122) = 2.26$, $p = .026$, $d = .40$). Although random assignment was used, the control classrooms had more students for whom English was a second language. Covariates were used to partially control for this flaw. Reliability for the CogAT was acceptable and consistent with previously published reliability estimates. Reliability could not be calculated for the fluency measure or the SAT-10 because item level data were not available.

3.2. Reading time

3.2.1. Adjusting times

Very short and very long reading times were removed or adjusted. Very short reading times can be caused by skimming as well as skipping clauses. To retain clauses that were skimmed but remove clauses that were skipped, we used the very conservative criterion of 400 ms as a minimum reading time. Since previous research has shown that college age students take at least 400 ms to read a word and press the space bar (Rayner & Clifton, 2009), elementary age students who read a clause and pressed the space bar in 400 ms or less probably did not read even one word; therefore these reading times were removed. Most students who had extremely short reading times had a very small number of them ($M = 1.9$; $Mdn = 0$), probably caused by keying errors, but 16 students had more than 10% of their reading times below 400 ms and were dropped from the analysis (6 reading for the purpose of preparing for a CR discussion, 3 reading to prepare for a regular discussion, 7 reading for no stated purpose but having participated in CR, and 4 reading for no stated purpose and not experiencing CR). Of the remaining students, the reading times for 154 clauses (1.1% of all clauses) were removed.

Any reading time that was more than a critical value of five standard deviations above the mean was replaced with the critical value. This conservative standard was used to take into account moments for reflection or the difficulties children may sometimes experience during reading. Eighty reading times (0.66% of the total cells) were trimmed. Considering removal of short times and

Table 1
Intercorrelations among measured variables.

	Pretest measures			Demographics		Treatments		Posttest measures	
	Fluency	CogAT	SAT-10	Age (months)	Gender	Collaborative reasoning	Announced discussion	SVT text comp.	Average reading time
Fluency	–	.38**	.65**	.16	−.20*	−.01	−.08	.20*	−.33**
CogAT		–	.46**	.20*	.07	.07	−.11	.21*	−.08
SAT-10			–	.03	−.06	.22*	.03	.23*	−.17
Age				–	.04	.14	−.17	.06	.03
Gender					–	.03	.28	.07	.24*
CR						–	.01	.03	−.18
AN							–	.00	.04
SVT								–	−.05
ART									–
M	65.55	8.75	53.72	116.66	–	–	–	1.07	3496.92
SD	14.32	3.95	19.36	6.27	–	–	–	.24	1195.62
Reliability	–	.83	–	–	–	–	–	.68	–

Note: SVT = SVT text comprehension test; ART = average reading time. CR and AN variables are effects coded so that positive one was given to students who had the collaborative reasoning treatment and the announcement treatment, and negative one was given to students who had regular discussions and no announcement.
*$p < .05$, **$p < .01$.

Table 2
Intercorrelations between clause-level variables.

	Syllables	Frequency	Propositions	Serial order	New arg.	Arg. wrap-up	Arg. order	After picture
Syllables	–	−.44**	.72**	−.01	.27**	−.12	−.20*	.13
Frequency		–	.222*	−.17	−.018	−.29	.127	−.10
Propositions			–	−.01	.15	.01	−.01	.12
Serial order				–	−.01	−.01	.02	−.02
New arg.					–	−.01	−.38**	.12
Arg. wrap-Up						–	.23*	−.11
Arg. order							–	−.03
After picture								–
M	9.43	9.17	3.78		.14	.11	4.29	.053
SD	3.79	.31	1.43		.38	.31	3.36	.23

Note: $^{*}p < .05$, $^{**}p < .01$.

Table 3
Reading time mixed effects model results.

	Model 1: Empty model	Model 2: Demographics	Model 3: Pretests	Model 4: Clause variables	Model 5: Treatments
Intercept	7.938*** (.04)	5.224** (.38)	6.814** (.36)	6.361** (.36)	6.479** (.37)
Age		.023** (.00)	.016** (.00)	.017** (.00)	.015** (.00)
Gender		.116 (.06)			
Fluency			−.001** (.00)	−.001** (.00)	−.001** (.00)
Number of propositions				.133** (.01)	.133** (.01)
Serial order				−.002** (.00)	−.002** (.00)
New argument				.095** (.02)	.095** (.02)
Collaborative reasoning					.073**(.02)
Announced discussion					.025 (.03)
Collaborative reasoning announced discussion interaction					.075** (.02)
Percent reduction in variance from Model 1		.70	1.71	5.92	6.37
AICc	15362.3	15317.9	14860.2	14106.5	14103.3

Note: $^{*}p < .05$, $^{**}p < .01$.

trimming of long times, there were a total of 234 altered times or 1.7% of the total number of recorded times. Clause reading times were then log transformed to correct for non-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 show that the natural log of the values closely followed the expected values.

3.2.2. Included variables

Student-level variables included reading fluency, spatial reasoning, reading comprehension, age, gender, discussion type (CR or regular), whether or not a future discussion was announced, and a set of dummy codes representing the student's model of the Earth before reading the text. A number of clause-level variables were entered to account for the variance explained by processing the reading at the three levels of representation in the construction–integration model. The number of syllables in the clause is a surface feature. The frequency that the word appears in the English language is a predictor of how easy it is to construct the textbase.[3] The number of propositions per clause is a representation of the information density of the clause and also reflects the time it takes to construct the textbase. The serial order, new argument, argument wrap-up, and argument order are variables representing the time it takes to construct the situation model. They are the order of the sentences from beginning to end, whether a clause begins a new argument from one of the girls in the story, whether a clause ends an argument, and the order of clauses from the beginning of an argument until the end.[4] An additional variable marked whether a clause appeared immediately after a picture to control for any variation due to the insertion of pictures into the story. The variables and their correlations are reported in Table 2.

3.2.3. Reading time model selection

The results of the reading time analysis are reported in Table 3. Model 1 is the empty model (ICC at the clause level = .243; ICC at the student level = .230). The subsequent models each improve model fit and explain additional variance. The final model (Model 5) consists of the clause-level variables number of propositions, new argument, and serial position; and the student-level variables fluency, discussion type, announcement of discussion, and the discussion type by announcement interaction (ICC at the clause level = .107, ICC at the student level = .275). All the covariates in the final model had significant effects in the expected directions.

The final model shows students who had CR discussions read more slowly and they read especially slowly if they received an announcement of a future discussion. It would not be meaningful to translate these standardized parameter estimate back into seconds; however, we can get a feeling for the effect size by observing that the standardized parameter estimate of the CR effect is about one half the standardized parameter estimate for the effect of propositions per clause. Therefore, reading for a CR discussion has about

[3] Frequency is how common a word is in the language. Since the participants were children, the frequencies were calculated from the Children's Printed Word Database which is a database of over 5000 children's books (Masterson, Stuart, Dixon, and Lovejoy, 2010).

[4] Propositions per clause was calculated using CPIDER Version 3.2 with speech mode turned off (Brown, Snodgrass, Kemper, Herman, & Covington, 2008). New argument, argument wrap up, and argument order were, respectively, the first clause in each argument, the last clause in each argument, and the sequential number of a clause in an argument (Britton, 1994).

Table 4
Logistic regression predicting conceptual growth on the last interview question from reading time residuals.

Independent variable		β	S.E.	Wald	df	Sig.	H–L test
Non-spherical clauses	Reading time Residuals	3.170	1.446	4.804	1	.028	7.195
	Constant	−1.358	.391	12.084	1	.001	
Spherical clauses	Reading time Residuals	−.011	.973	.000	1	.991	10.980
	Constant	−1.073	.321	11.141	1	.001	

one half the effect of adding one standard deviation in the number of propositions for each clause. When this effect is added over 110 clauses it amounts to a similar amount of time the average participant needed to read one page of the six page text or slightly less than one hundred words.

3.3. Analysis of the Sentence Verification Test of comprehension

The main purpose of SVT comprehension test was to evaluate whether increased reading times were caused by distracted reading rather than increased engagement. The SVT data had a negative skew partly due to a performance ceiling (*Mdn* = 11 out of a possible 15). The SVT was square-root arcsine transformed to adjust for the negative skew. The transformed data met the Kolmogorov–Smirnov test for a normal distribution ($p = .25$). Nine students who were learning English as a second language were removed because their low comprehension scores were likely to have been strongly influenced by factors other than reading speed, such as vocabulary knowledge (two students reading for the purpose of preparing for a CR discussion, three students reading for the purpose of preparing for a regular discussion, one student reading without a stated purpose but who did experience CR, and three students reading without a stated purpose who did not experience CR).

If increased reading time was caused by increased cognitive engagement with the text rather than distraction, it would be expected that students' reading times should predict SVT comprehension score after controlling for students general SAT-10 reading comprehension scores in a regression framework, which is the result that was found ($\beta = .271$; $p = .028$).

3.4. Analyses of the shape-of-the-Earth interview

Logistic regressions were used to predict conceptual growth on the last question from the reading times on a selected subset of clauses. Two subsets were used. One subset of 16 clauses uttered by the character arguing for a spherical Earth and one subset of 20 clauses uttered by the character arguing for against a spherical Earth.

Model 4 of the reading time data (Section 3.2) was run and studentized residuals were calculated for each clause. These residuals represent the amount that the student deviated from the time that would be expected given the characteristics of the clause and the student's general reading fluency without the treatments. The average of the residuals for each subset of clauses was entered into the logistic regression as the independent variable. The dependent variable was whether students progressed on the final interview question. Since only students who had a pretest answer that was not consistent with a spherical model of the Earth had an opportunity to progress only 67 participants were used in this analysis. To compare these students to the students who began with an answer consistent with a spherical model of the Earth on the last question, the residuals for each subset of clauses was calculated.

3.4.1. Results of the logistic regression predicting conceptual growth

The Hosmer–Lemeshow test was insignificant for both types of clauses showing that progress was well modeled by reading time residuals (Table 4). The results show that reading time residuals for the clauses arguing for a spherical Earth did not predict conceptual growth whereas reading time residuals for the clauses arguing against a spherical Earth did predict conceptual growth (a one standard deviation in the studentized residuals resulted in a 23.8 times increase in the odds of improving on the last question).

3.4.2. Comparison between reading times from those who began with and without an answer consistent with a spherical model of the Earth

To further explore the significant results found in the logistic regression the means of the residuals were analyzed. Students who began with an answer consistent with a spherical model of the Earth tended to read both clauses that argue for and clauses that argued against a spherical Earth close to the speed that would be predicted by the model (mean residual for spherical clauses = .050, mean residual for non-spherical clauses = .034, 0 indicates a time equal to the predicted time). Students who began with a non-spherical answer and did not change also read both sets of clauses at close to the predicted speed (mean residual for spherical clauses = .040, mean residual for non-spherical clauses = −.009). In contrast students who began with a non-spherical answer and did change to a spherical model, read clauses arguing for a spherical Earth more quickly than would have been predicted and clauses arguing against a spherical Earth more slowly than would have been predicted (mean residual for spherical clauses = −.163, mean residual for non-spherical clauses = .164). This suggests that students who improved their models did so more by reevaluating their prior beliefs than by reconsidering the scientific conception.

3.4.3. Analyzing the overall effect of the discussion purpose for reading on conceptual change

A logistic regression was used with demographic status, ability measures, and students' initial model of the earth as covariates. There were 102 usable pre-test interviews and 96 usable post-test interviews. The distribution of models identified in this sample (Table 5) is similar to those of Vosniadou and Brewer (1992), but a larger percentage of students have mixed models. This might be due to the slightly different interview procedures (see Section 2.3.2).

3.4.4. Change in models

To be included in this analysis, students had to have complete pre- and posttest interviews and they needed to begin the study with something other than a spherical model of the Earth

Table 5
Frequency of models from shape-of-the-Earth interview.

Posttest model	Sphere strict	Sphere loose	Hollow strict	Hollow loose	Dual	Flattened loose	Disc loose	Mixed	Total
Pretest	18	23	7	8	2	1	2	38	102
Posttest	44	15	11	2	1	1	2	19	96

Table 6
Conceptual change analysis by treatment.

Pretest model	Posttest model				
	Strict sphere	Loose sphere	Strict hollow	Loose hollow	Mixed
Reading for discussion/participated in CR					
Strict sphere	**5**				
Loose sphere	5	**3**			1
Strict hollow	1				1
Loose hollow					
Mixed	4	3	1	2	**2**
Total	15	6	1	2	4
Reading for discussion/no experience with CR					
Strict sphere					
Loose sphere	2				
Strict hollow					1
Loose hollow					
Mixed	1	2	1		**3**
Total	3	2	1		4
No stated purpose for reading/participated in CR					
Strict sphere	**7**			1	
Loose sphere	5				1
Strict hollow					1
Loose hollow				**1**	
Mixed	3	1	1		
Total	15	1	1	2	**2**
No stated purpose for reading/no experience with CR					
Strict sphere	**3**	2	1		
Loose sphere	3				
Strict hollow	1				
Loose hollow			1		1
Mixed	2	1		1	**2**
Total	9	3	2	1	3
Grand total	42	12	5	5	13

Note. The models are listed in Brewer's developmental order (2008) with each model subdivided into strict and loose consistency. Only models with more than two entries were included for readability. Students who had no change are in bold.

(otherwise it would be impossible to show improvement). There were 83 students who met these criteria. Table 6 presents the results according to treatment. Due to differences in absences, class size, and other causes, there were differences in the number of students who had complete sets of interviews. Of the 28 students who read to prepare for a CR discussion, 16 progressed. Of the 10 students who read for a regular discussion, 6 progressed. Of the 21 students who did not read for a stated purpose but did have CR experiences, 10 progressed. Of the 18 students who did not read for a stated purpose and did not experience CR discussion, 8 progressed. The majority of these changes was from a loose to a strict spherical model or from mixed to a variety of other models.

3.4.5. *Shape-of-the-Earth model selection*

Several covariates were used in this analysis including age, gender, the CogAt spatial reasoning score, and two measures of students' pretest models. The CogAt score was entered, because it was thought that creating and modifying one's model of the Earth inherently involves manipulating objects in the mind. Two variables were entered to control for students' initial conceptions: sphere indicates whether a student did or did not have a loose spherical model of the Earth at the time of the pretest, and mixed indicates if a student had a mixed or coherent model at the time of the pretest.

As shown in Table 7, Model 2 shows that higher scores on the spatial reasoning test and a mixed pretest model strongly predicted a student's probability of progress in understanding the shape of the Earth. Students who progressed had spatial reasoning scores almost one standard deviation above students who did not ($M = 9.60$, $SD = 3.37$; $M = 6.38$, $SD = 3.72$ respectively). Students who began with a mixed model had odds of progressing 22% higher than students who began with a coherent model (23/31 or 76%; 21/39 or 54%, respectively).

The final model, Model 4, includes the treatment variables and the interactions between the treatment conditions and whether students had a mixed pretest model. The Hosmer–Lemeshow test as well as the Cox & Snell R^2 show that this model has a good fit to the data. Model 4 produced an improved model fit and an increase in R^2 compared to Model 2. The saturated model, Model 5, was an even better fit, but the very large standard errors suggest that the model is overfitted.

Model 4 indicate that after controlling for spatial reasoning ability, reading for the purpose of preparing for discussion did not increase the rate of conceptual change. However, participating in CR discussions did increase the rate of conceptual change for some students (Fig. 2). Participating in CR discussions increased conceptual change for students who started with a mixed model. It increased the odds of progressing by 5.63 times (controlling for spatial reasoning ability). For students who began with a coherent model, the odds of progressing actually decreased by .67 times (controlling for spatial reasoning ability). This interaction suggests

Table 7
Conceptual change model results.

	Model 1: Demographics	Model 2: Pretests	Model 3: Treatments	Model 4: Interactions[a]	Model 5: Saturated
Intercept	−2.86 (4.38)	−1.32 (1.60)	−2.45* (.94)	−2.90** (1.01)	−3.60** (1.32)
Age	.024 (.04)				
Gender	.27 (.49)				
CogAt		.22* (.09)	.18** (.08)	.23** (.09)	.408** (.126)
Sphere model		−.186 (.85)			
Mixed model		2.07* (.86)	2.20** (.68)	2.66** (.79)	6.91 (4510.9)
Collaborative reasoning			−.09 (.28)	−.61 (.35)	−.368 (.475)
Announced discussion			−.14 (.28)	−.02 (.33)	−.187 (.475)
Collaborative reasoning*Announced discussion interaction			.03 (.28)	−.20 (.32)	−.041 (4510.9)
Collaborative reasoning*Mixed model interaction				1.61** (.59)	−4.75 (4510.9)
Announced discussion*Mixed model interaction				−.68 (.59)	−4.86 (4510.9)
Collaborative reasoning*Announced discussion*Mixed model					−4.75 (4510.9)
Hosmer and Lemeshow Test	4.58	5.82	7.05	11.10	14.3
Cox & Snell R^2	.015	.234	.215	.289	.353
AICc	102.55	90.24	89.88	91.09	75.71

Note: $^*p < .05$, $^{**}p < .01$.

[a] Model 4 is the selected model. It assumes that mixed models are less sophisticated than hollow models. If mixed models are assumed to be more sophisticated than hollow models the estimates would be as follows: intercept = −3.66** (1.254); CogAt = .469** (.127); Mixed = −1.077 (.799); CR = .756 (.974); Announced = .312 (.950); CR*Announced = −.139 (.366); Cr*Mixed = 1.46* (.719); Announced*Mixed = .266 (.718).

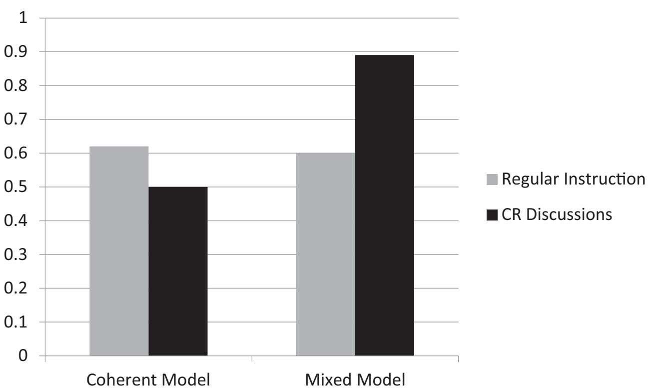

Fig. 2. Odds of conceptual progress from pre-test to post-test as a function of type of pre-test model and type of discussion.

that the experience of argumentative discussion led students who already had a coherent model to become slightly more entrenched in their beliefs, but it strongly induced students who started with an incoherent model of the shape of the Earth to develop a model closer to the spherical model.

4. Discussion

This experiment tested the effect of classroom discussion as a purpose for reading science texts. The Cognitive Reconstruction of Knowledge Model (Dole & Sinatra, 1998) was used as a theoretical framework. The CRKM predicts that warm variables can increase conceptual change through the mechanism of deeper engagement. This mediation model was divided into three hypotheses. Hypothesis 1 was directed at the first path of the mediation model in which purpose of reading explained reading time variance after accounting for other factors. This hypothesis was supported by the reading time results. The alternative explanation that the slower reading times might have been caused by students being distracted by thinking about the upcoming discussion was countered by the result that students who read more slowly also had better comprehension of the text.

The second path in the mediation model leads from reading times to conceptual change. This path was analyzed for a subset of the reading times and interview questions. These results did show a significant path but only for reading times for clauses arguing against the spherical shape of the Earth. Comparing those students who progressed and students who did not progress or who began with a spherical Earth model demonstrated that students who progressed showed an allocation pattern in which they preferentially spent their cognitive resources on clauses discussing their current conception. This suggests that these students progressed because they spent extra time reevaluating their prior beliefs. This pattern of allocation also suggests that these students might have already believed in the spherical Earth, but also had misconceptions that they had not discarded. By reading their misconceptions in full they might have been able to finally reevaluate these issues so that they could fully embrace the spherical model. This conjecture is fully concordant with the warm conceptual change approach of the CRKM in which motivated students engage in metacognitive reflection.

The results of the test for Hypothesis 3 found no effect of announcing to students that they were reading for the purpose of preparing for a discussion, but there was an effect for participating in CR discussions under certain conditions. Students who participated in CR discussions had an increased rate of conceptual change if they began with a mixed model of the Earth; almost 90% of them adopted the scientific model. However, students who participated in CR discussions and started with a coherent but non-spherical model had less conceptual growth than their peers.

These two results can be at least partially explained using the well-established finding in persuasion research called *forewarning* (Benoit, 1998). According to this theory, people who are anticipating an attempt to persuade them to change a strongly held view are more likely to generate counterarguments and resist persuasion whereas people who are anticipating a message consistent with their views or who have weak views are receptive to strong arguments. This has also been shown by Nussbaum (2005) who found that students writing for the purpose of persuasion showed greater my-side bias.

As predicted by forewarning research, students with consistent views in this experiment became more entrenched in their beliefs if they anticipated an argumentative discussion. They likely anticipated receiving challenging arguments and made efforts to contradict those arguments when they appeared. Since the story was written as a two-sided persuasive text, all students received contrary arguments. This process of producing counterarguments might have made them even more certain of their prior beliefs. In contrast students with uncertain views who were anticipating a discussion were more receptive to strong arguments and more receptive to reexamining their previously held beliefs.

The results of the third analysis further supports the hypothesized mediated model but in a more contingent way. Only the CR treatment increased the rate of conceptual change and only for students with mixed pretest models. The results of the third analysis emphasize that warm conceptual change theories are intended to expand rather than negate more traditional explanations of conceptual change. As the CRKM predicts, increased cognitive engagement does not always lead to conceptual change. It is more likely and when it occurs it is stronger, but it allows for the fact that people can think more about a topic and yet still not be convinced by the evidence. This experiment suggests some of the circumstances that might contribute to that very result including exposing students to scientific theories before they are unsatisfied with their strongly held prior conceptions. Furthermore, even when students are prepared for change, their increased level of cognitive engagement can be made more productive by exposing them to argumentative discussions and framing lessons around an argumentative format such as the CR approach. It is not surprising that this approach would prompt students to use their cognitive resources more productively given all the previous findings regarding the transfer of argumentative discussion to students individual thinking (Reznitskya et al., 2009). Reznitskya et al. explains these results in terms of a combination of schema theory and socio-cognitive theory; from participating in CR discussions students can internalize an argument schema that allows them to form arguments with more reasons, counterarguments, and rebuttals. This would help students who were already unsure about their position to reconsider their arguments and form a new more coherent conception, while it would also allow students with strong misconceptions to be somewhat more skilled at answering counterarguments with rebuttals. Reading to prepare for a future discussion might cause students to read more carefully, but it would not necessarily give students these tools. Both the 'warm' variables of increased engagement and the 'cold' variables of improved argumentation skills are needed.

4.1. Dual-process theories

The observed results demonstrate the usefulness of dual-process theories and particularly elements of the CRKM. As predicted by the dual-processing approach, anticipating discussion led to increased cognitive engagement especially for argumentative discussions. Furthermore, increased engagement led to increased conceptual growth. Conceptual change overall had uneven results despite increased engagement. Only students who began with incoherent models showed increased rates of conceptual change. This result is consistent with the CRKM as the coherence in learner's existing conceptions is part of the model. While the present study was not a test of the CRKM theory as a whole, it does demonstrate the theory's value in creating hypotheses, inspiring research designs, and helping to interpret results. This study also expanded the use of the CRKM to elementary age children and reveals some of the complexities of investigating the interacting variables of the theory.

4.2. Study limitations and areas for future research

One of the primary limitations of this study included the lack of a multifaceted assessment of engagement. Reading time is a strong measure of cognitive engagement but is not a typically used measure of emotional and behavioral engagement. Future research needs to expand findings on the warm effects of discussion by seeking to measure the emotional and behavioral changes associated with discussion. Once again, isolating these effects from the effects of sharing information and co-constructing meaning through discourse will be challenging.

Another limitation of the study is that this experiment only studied two sets of discussion expectations, which makes it impossible to identify which of the many factors associated with each set of expectations was important for increasing engagement and subsequent conceptual change. For example, Felton, Garcia-Mila, and Gilbert (2009) divide argumentative discourse goals into three types: non-argumentative, deliberative, and disputative. Dispute discussions have as a goal to persuade others while deliberative discussion has the goal of finding consensus. Felton, Garcia-Mila, and Gilbert found that while students participating in both argumentative discussions had greater learning gains than control discussions, students who participated in deliberative discussions outperformed students who participated in disputative discussions. CR is a discussion approach that incorporates aspects of both dispute and deliberation. A CR discussion has a dispute goal in that students are encouraged to take a position supported by reasons and evidence, but it also has a deliberative goal in that students are encouraged to consider both sides of every issue. In the future, more specific discussion manipulations can be used to identify what about discussions is most important in increasing engagement and how teachers can foster these aspects.

An associated line of research is also needed in terms of what type of text is most productive. Much research in recent years has been focused on refutational texts that strongly argue for one side of the issue, but this experiment used a two-sided text framed as a conversation with two individuals arguing for their own side. This might have led to some students becoming entrenched in their views. On the other hand, it did lead to substantial conceptual change among some students despite previous research suggested that two-sided nonrefutational texts are not persuasive (Allen, 1991; Murphy, 2001). Understanding the details of the interaction between persuasive text characteristics and student characteristics will be important in deciding on how to best design texts that maximize the refutational text effect.

Other limitations were not specific to this study because they are inherent in the current methodologies and technologies available to measure engagement. The experiment was designed to analyze conceptual growth over a brief reading task. While work at small grain sizes is critical to understanding the basic mechanisms of psychological processes, ultimately the theory is intended to explain lessons that contain iterative cycles of engagement including lessons following an inquiry format. Highly sensitive online methods like reading time need to be developed that can function over extended time periods to allow them to be used during more authentic learning activities. This would also allow for effect sizes large enough to test for persistence effects over delayed post testing, and it would allow for a more fine-grained analysis of conceptual change than achieved in this experiment.

4.3. Creating a context of socially accountable discourse

The fact that reading for the purpose of preparing for a discussion and especially for CR discussions had a measurable impact on how students read a text sends an important message about the power of creating a context of socially accountable discourse. As Michaels, O'Connor, and Resnick (2008) describe, Accountable Talk is not just a method of creating a certain exchange of useful words, it creates a classroom climate that changes the entire experience of education. The present study shows that these effects are present even when students are completely silent. Certainly, argumentative discussions are important in supporting critical thinking or sharing information, but it also changes the way students feel and think about the entire learning experience. This suggests that if instruction was framed within a context of argumentative discussion, it could transform the classroom climate creating an enduring culture of thoughtfulness.

Acknowledgments

First and foremost we would like to thank the participating teachers and students. We would also like to thank Aini Marina Ma'rof, Kim Nguyen-Jahiel and the other individuals who helped with data collection and transcription. We would like to thank Dr. Jennifer Cromley for her very helpful feedback on earlier drafts of this manuscript.

We would also like to thank the Institute of Education Sciences, U.S. Department of Education, who provided partial funding for this project through Grant R305A080347 to the Board of Trustees of the University of Illinois. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. They did not contribute to the project other than providing funding.

References

Abd-El-Khalick, F., & Akerson, V. L. (2004). Learning as conceptual change: factors mediating the development of preservice elementary teachers' views of nature of science. *Science Teacher Education, 88*, 785–810. http://dx.doi.org/10.1002/sce.10143.

Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: John Wiley & Sons.

Alexander, P. A., Fives, H., Buehl, M. M., & Mulhern, J. (2002). Teaching as persuasion. *Teaching and Teacher Education, 18*, 795–813. http://dx.doi.org/10.1016/S0742-051X(02)00044-6.

Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication, 55*, 390–404. http://dx.doi.org/10.1080/10570319109374395.

Anderson, R. C., Chinn, C., Chang, J., Waggoner, J., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. L. Osborn (Ed.), *Literacy for all: Issues in teaching and learning* (pp. 170–186). New York: Guilford.

Au, K. H.-P. (1979). Using the experience-text-relationship method with minority children. *The Reading Teacher, 32*, 677–679.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the Author: a yearlong classroom implementation to engage students with text. *The Elementary School Journal, 96*, 385–414.

Bennett, J., Hogarth, S., Lubben, F., Campbell, B., & Robinson, A. (2010). Talking Science: the research evidence on the use of small group discussions in science teaching. *International Journal of Science Education, 32*, 69–95. http://dx.doi.org/10.1080/09500690802713507.

Benoit, W. L. (1998). Forewarning and persuasion. In M. Allen, & R. Priess (Eds.), *Persuasion: Advances through meta-analysis* (pp. 159–184). Cresskill, NJ: Hampton Press.

Brewer, W. F. (2008). Naïve theories of observational astronomy: Review, analysis, and theoretical implications. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 155–204). New York: Routledge.

Britton, B. K. (1994). Understanding expository text: Building mental structures to induce insights. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). New York, NY: Academic Press.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22*, 335–351. http://dx.doi.org/10.1002/acp.1418.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*, 1081–1087. http://dx.doi.org/10.3758/BF03206376.

Broughton, S. H., Sinatra, G. M., & Nussbaum, E. M. (2011). "Pluto has been a planet my whole life!" Emotions, attitudes, and conceptual change in elementary students learning about Pluto's reclassification. *Research in Science Education, 43*, 529–550. http://dx.doi.org/10.1007/s11165-011-9274-x.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40*, 540–545. http://dx.doi.org/10.3758/BRM.40.2.540.

Cazden, C. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752–766. http://dx.doi.org/10.1037/0022-3514.39.5.752.

Chi, M. T. H. (2008). Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research in conceptual change* (pp. 61–82). New York, NY: Routledge.

Chinn, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly, 36*, 378–411. http://dx.doi.org/10.1598/RRQ.36.4.3.

Clark, A., Anderson, R. C., Kuo, L., Kim, I., Archodidou, A., & Nguyen-Jahiel, K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review, 15*, 181–198. http://dx.doi.org/10.1023/A:1023429215151.

Clause. (2013). *Oxford English Dictionary online*. Retrieved from http://www.oed.com/view/Entry/33948?rskey=4rVJCF&result=1#eid.

Cobern, W. W. (1996). Worldview theory and conceptual change in science education. *Science & Education, 4*, 287–302.

Craik, F. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*, 109–128. http://dx.doi.org/10.1080/00461520.1998.9653294.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of argumentation in classrooms. *Science Education, 84*, 287–312. http://dx.doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446. http://dx.doi.org/10.1016/0749-596X(88)90066-6.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671–688. http://dx.doi.org/10.1080/09500690305016.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 75–87. http://dx.doi.org/10.1016/S0022-5371(83)80007-3.

Felton, M., Garcia-Mila, M., & Gilabert, S. (2009). Deliberation versus Dispute: the impact of argumentative discourse goals on learning and reasoning in the science classroom. *Informal Logic, 4*, 417–446.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109. http://dx.doi.org/10.3102/00346543074001059.

Goldenberg, C. (1992). Instruction conversations: promoting comprehension through discussion. *The Reading Teacher, 46*, 316–326.

Greene, B., Miller, R., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: contributions of classroom perceptions and motivation. *Contemporary Educational Psychology, 29*, 462–482. http://dx.doi.org/10.1016/j.cedpsych.2004.01.006.

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review, 15*, 147–179. http://dx.doi.org/10.1023/A:1023477131081.

Hayes, B. K., Goodhew, A., Heit, E., & Gillan, J. (2003). The role of diverse instruction in conceptual change. *Journal of Experimental Child Psychology, 86*, 253–276. http://dx.doi.org/10.1016/j.jecp.2003.09.002.

He, Y. (2007). *Influence of reading proficiency on text representation in L1 and L2* (Doctoral dissertation). Retrieved June 6, 2012 from http://vufind.carli.illinois.edu/vf-uiu/Record/uiu_5762659/.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: foundations for the twenty-first century. *Science Education, 88*, 28–54. http://dx.doi.org/10.1002/sce.10106.

Hynd, C. (2003). Conceptual change in response to persuasive message. In G. M. Sinatra, & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 291–316). Mahwah, New Jersey: Lawrence Erlbaum.

Jacobs, J. (1902). *The fables of Aesop*. New York, NY: Macmillan.

Jimenez-Aleixandre, M. P., & Erduran, S. (2008). Argumentation in science education: an overview. In S. Erduran, & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–27). Dordrecht, Netherlands: Springer.

Jiménez-Aleixandre, M. P., Rodríguez, A. B., & Duschl, R. (2000). "Doing the lesson" or "doing science": argument in high school genetics. *Science Education, 84*, 757–792. http://dx.doi.org/10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F.

Johnson, B. T., & Eagly, A. H. (1989). Effects of involvement on persuasion: a meta-analysis. *Psychology Bulletin, 106*, 290–314. http://dx.doi.org/10.1037//0033-2909.106.2.290.

Johnson, M. L., & Sinatra, G. M. (2013). Use of task-value instructional inductions for facilitating engagement and conceptual change. *Contemporary Educational Psychology, 38*, 51–63. http://dx.doi.org/10.1016/j.cedpsych.2012.09.003.

Kaakinen, J. K., & Hyona, J. (2005). Perspective effects on expository text comprehension: evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes, 40*, 239–257. http://dx.doi.org/10.1207/s15326950dp4003_4.

Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education, 21*, 431–446. http://dx.doi.org/10.1080/095006999290642.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95*, 163–182.

Kuhn, D. (1993). Science as argument: implications for teaching and learning scientific thinking. *Science Education, 77*, 319–337. http://dx.doi.org/10.1002/sce.3730770306.

Lelliott, A., & Rollnick, M. (2009). Big ideas: a review of astronomy education research. *International Journal of Science Education, 32*, 1771–1799. http://dx.doi.org/10.1080/09500690903214546.

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, New Jersey: Ablex Publishing.

Linnenbrink, E. A., & Pintrich, P. R. (2002). The role of motivational beliefs in conceptual change. In M. Limon, & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 29–60). Norwell, MA: Kluwer Academic Publishers.

Lohman, D. F. (2003a). *The Wechsler Intelligence Scale for Children III and the Cognitive Abilities Test (Form 6): Are the general factors the same?*. Retrieved December 12, 2011, fromhttp://faculty.education.uiowa.edu/dlohman.

Lohman, D. F. (2003b). *The Woodcock-Johnson III and the Cognitive Abilities Test (Form 6): A concurrent validity study*. Retrieved December 12, 2011, fromhttp://faculty.education.uiowa.edu/dlohman.

Lombardi, D., Sinatra, G. M., & Nussbaum, E. M. (2013). Plausibility reappraisals and shifts in middle school students' climate change conceptions. *Learning and Instruction, 27*, 50–62. http://dx.doi.org/10.1016/j.learninstruc.2013.03.001.

Mansfield, A., & Clinchy, B. (2002). Toward the integration of objectivity and subjectivity: epistemological development from 10 to 16. *New Ideas in Psychology, 20*, 225–262. http://dx.doi.org/10.1016/S0732-118X(02)00008-9.

Mason, L. (2001). The process of change through persuasion: a commentary. *International Journal of Educational Research, 35*, 715–729. http://dx.doi.org/10.1016/S0883-0355(02)00011-3.

Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: the interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology, 100*, 291–309. http://dx.doi.org/10.1037/0022-0663.100.2.291.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology, 101*, 221–242. http://dx.doi.org/10.1348/000712608X371744.

McKeown, M. G., Beck, I. L., & Worthy, M. J. (1993). Grappling with text ideas: questioning the author. *The Reading Teacher, 46*, 560–5613.

Mehan, H. (1979). "What time is it, Denise?": asking known information questions in classroom discourse. *Theory into Practice, 18*, 285–294. http://dx.doi.org/10.1080/00405847909542846.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*, 283–297.

Mikkilä-Erdmann, M., Penttinen, M., Anto, E., & Okinuora, E. (2008). Constructing mental models during learning from science text. In P. Ifenthaler, P. Pimay-Dummer, & J. M. Spector (Eds.), *Understanding models for learning and instruction: Essays in honor of Norbert M. Steel* (pp. 62–77). New York: Springer.

Miller, B. (2009). *What shape is the Earth?* Champaign, IL: Center for the Study of Reading.

Murphy, P. K. (2001). What makes a text persuasive? Comparing students' and experts' conceptions of persuasiveness. *International Journal of Educational Research, 35*, 675–698. http://dx.doi.org/10.1016/S0883-0355(02)00009-5.

Murphy, P. K. (2007). The eye of the beholder: the interplay of social and cognitive components in change. *Educational Psychologist, 42*, 41–53. http://dx.doi.org/10.1080/00461520709336917.
Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: a meta-analysis. *Journal of Educational Psychology, 101*, 740–764. http://dx.doi.org/10.1037/a0015576.
Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30*, 286–313. http://dx.doi.org/10.1016/j.cedpsych.2004.11.002.
Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384–395. http://dx.doi.org/10.1016/S0361-476X(02)00038-3.
O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 410–420. http://dx.doi.org/10.1037/0278-7393.14.3.410.
Osborne, J. (2010). Arguing to learn in science. The role of collaborative, critical discourse. *Science, 328*, 463–466. http://dx.doi.org/10.1126/science.1183944.
Palinscar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175.
Pearson Inc.. (2003). *Stanford English Language Proficiency test technical manual*. San Antonio, TX: Pearson Inc.
Pedhazur, E. (1997). *Multiple regression in behavioral research*. Orlando, FL: Harcourt Brace.
Penttinen, M., Anto, E., & Mikkilä-Erdmann, M. (2013). Conceptual change, text comprehension and eye movements during reading. *Research in Science Education, 43*, 1407–1434. http://dx.doi.org/10.1007/s11165-012-9313-2.
Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routs to attitude change*. New York: Springer-Verlag.
Petty, R. E., Rucker, D. D., Bizer, G. Y., & Cacioppo, J. T. (2004). The elaboration likelihood model of persuasion. In J. S. Seiter, & R. H. Gass (Eds.), *Perspectives on persuasion, social influence, and compliance gaining* (pp. 65–89). Boston: Pearson Education.
Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167–199. http://dx.doi.org/10.3102/00346543063002167.
Pontecorvo, C. (1987). Discussing and reasoning: the role of argument in knowledge construction. In E. De Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction: European research in an international context* (pp. 239–250). Oxford: Pergamon Press.
Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education, 66*, 211–227. http://dx.doi.org/10.1002/sce.3730660207.
Ranellucci, J., Muis, K. R., Duffy, M., Wang, X., Sampasivam, L., & Franco, G. M. (2012). To master or perform? Exploring relations between achievement goals and conceptual change learning. *British Journal of Educational Psychology, 83*, 431–451. http://dx.doi.org/10.1111/j.2044-8279.2012.02072.x.
Raphael, T., & McMahon, S. (1994). Book club: an alternative framework for reading instruction. *The Reading Teacher, 48*, 102–116.
Rayner, K., & Clifton, C., Jr. (2009). Language processing in reading and speech perception is fast and incremental: implications for event potential related research. *Biological Psychology, 80*, 4–9. http://dx.doi.org/10.1016/j.biopsycho.2008.05.002.
Reynolds, R. E., & Anderson, R. C. (1982). Influence of questions on allocation of attention during reading. *Journal of Educational Psychology, 74*, 623–632.
Reznitskaya, A., Kuo, L.-J., Clark, A.-M., Miller, B., Jadallah, M., Anderson, R. C., et al. (2009). Collaborative reasoning: a dialogic approach to group discussions. *Cambridge Journal of Education, 39*, 29–48.
Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes, 41*, 221–250. http://dx.doi.org/10.1207/s15326950dp4103_1.
Riverside Publishing. (2001). *Cognitive Abilities Test (CogAT), Form 6*. Rolling Meadows, IL: Riverside Publishing.
Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: a review of nineteen experimental studies. *Review of Educational Research, 64*, 479–530.
Rothkopf, E. Z., & Billington, M. J. (1979). Goal guided learning from text: inferring a descriptive processing model from inspection times and eye movements. *Journal of Educational Psychology, 71*, 310–327.
Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction, 14*, 373–408. http://dx.doi.org/10.1207/s1532690xci1403_4.
Sharp, A. M. (1995). Philosophy for children and the development of ethical values. *Early Child Development and Care, 197*, 45–55. http://dx.doi.org/10.1080/0300443951070106.
Shepard, R. N., & Feng, C. A. (1972). A chronometric study of mental paper folding. *Cognitive Psychology, 3*, 228–243. http://dx.doi.org/10.1016/0010-0285(72)90005-9.
Sinatra, G. M. (2005). The "warming trend" in conceptual change research: the legacy of Paul R. Pintrich. *Educational Psychologist, 40*, 107–115. http://dx.doi.org/10.1207/s15326985ep4002_5.
Sinatra, G. M., & Heddy, B. C. (2013). *The challenges of defining and measuring student engagement in science learning*. Paper presented at The 2013 Annual Meeting of the American Educational Research Association: Education and Poverty: Theory, Research, Policy and Praxis, San Franciso, CA.
Sneider, C., & Pulos, S. (1983). Children's cosmographies: understanding the earth's shape and gravity. *Science Education, 67*, 205–221. http://dx.doi.org/10.1002/sce.3730670209.
Solomon, J. (1993). *Teaching science, technology and society*. Buckingham, U.K.: Open University Press.
Spies, R. A., & Plake, B. S. (2005). *The sixteenth mental measurements yearbook*. Lincoln, NE: The University of Nebraska Press.
Taasoobshirazi, G., & Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching, 48*, 901–918. http://dx.doi.org/10.1002/tea.20434.
Vosniadou. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69. http://dx.doi.org/10.1016/0959-4752(94)90018-3.
Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: a study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585. http://dx.doi.org/10.1016/0010-0285(92)90018-W.
Wade, S. E., Schraw, G., Buston, W. M., & Hayes, M. T. (1993). Seduction of the strategic reader: effects of interest on strategies and recall. *Reading Research Quarterly, 28*, 93–114.
Wang, M. T., Willett, J. B., & Eccles, J. S. (2011). The assessment of school engagement: examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology, 49*, 465–480. http://dx.doi.org/10.1016/j.jsp.2011.04.001.
Woods, B. S., & Murphy, P. K. (2001). Separated at birth: the shared lineage of research on conceptual change and persuasion. *International Journal of Educational Research, 35*, 633–649. http://dx.doi.org/10.1016/S0883-0355(02)00007-1.
Wu, X., Anderson, R. C., Nguyen-Jahiel, K., & Miller, B. (2013). Enhancing motivation and engagement through collaborative discussion. *Journal of Educational Psychology, 105*, 622–632. http://dx.doi.org/10.1037/a0032792.
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*, 35–62. http://dx.doi.org/10.1002/tea.10008.