Learning and Instruction 27 (2013) 50–62

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

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

Plausibility reappraisals and shifts in middle school students' climate change conceptions

Doug Lombardi [a,*], Gale M. Sinatra [b], E. Michael Nussbaum [c]

[a] Temple University, 450 Ritter Hall, 1301 Cecil B. Moore Avenue, Philadelphia, PA 19122, USA
[b] University of Southern California, USA
[c] University of Nevada, Las Vegas, USA

ARTICLE INFO

Article history:
Received 15 June 2012
Received in revised form
1 March 2013
Accepted 9 March 2013

Keywords:
Plausibility perceptions
Critical evaluation
Conceptual change
Epistemic cognition
Climate change

ABSTRACT

Plausibility is a central but under-examined topic in conceptual change research. Climate change is an important socio-scientific topic; however, many view human-induced climate change as implausible. When learning about climate change, students need to make plausibility judgments but they may not be sufficiently critical or reflective. The purpose of this study was to examine how students' plausibility judgments and knowledge about human-induced climate change transform during instruction promoting critical evaluation. The results revealed that treatment group participants who engaged in critical evaluation experienced a significant shift in their plausibility judgments toward the scientifically accepted model of human-induced climate change. This shift was accompanied by significant conceptual change postinstruction that was maintained after a six-month delay. A comparison group who experienced a climate change activity that is part of their normal curriculum did not experience statistically significant changes.

1. Introduction

Climate change science is complex. Developing full understanding about Earth's climate requires fundamental knowledge in many domains. Many of these scientific ideas are counter to students' existing mental representations. These alternative or naïve mental representations form via experiences with the natural world, as well as experiences at school and other everyday interactions (see, for example, Vosniadou, 1994; Vosniadou & Brewer, 1992). However, many alternative conceptions are notoriously robust to change and can act as a barrier to learning scientifically accurate ideas (Chi, 2005).

Plausibility judgments may be an important way in which students evaluate scientific concepts to facilitate reconstruction of alternative knowledge structures into scientifically accurate conceptions. We define plausibility as a judgment on the *relative potential truthfulness* of incoming information compared to our existing mental representations (Lombardi, Nussbaum, & Sinatra, in review). In essence, plausibility judgments about ideas generally achieve a lesser standard than firmer epistemic commitments, such as judgments about what an individual believes to be correct (Rescher, 1976; Southerland, Sinatra, & Matthews, 2001). In reflecting on plausible reasoning, Rescher (1976) states that "the 'acceptance' of a proposition as a potential truth is not actual *acceptance* of it at all, but a highly provisional and conditional epistemic inclination towards it, an inclination that falls far short of outright commitment" (Rescher, 1976, p. 9, emphasis in original).

1.1. Plausibility judgments and conceptual change

Plausibility judgments can arise in situations of competing explanations. Conceptual change theorists often contend that scientifically accurate conceptions must first be judged as plausible—ultimately as more plausible than existing conceptions—for conceptual change to occur (Dole & Sinatra, 1998; Pintrich, Marx, & Boyle, 1993; Posner, Strike, Hewson, & Gertzog, 1982). The nature of plausibility judgments has not been sufficiently investigated, however, by conceptual change researchers. A recent theoretical model on the role of plausibility in conceptual change was developed by Lombardi et al. (in review) (Fig. 1). The structure of the model is based primarily on Rescher's (1976) model of plausible reasoning, as well on Connell and Keane's (2006) model of plausibility. To account for conceptual change however, the model draws heavily on aspects of Dole and Sinatra's (1998) CRKM model, which has influenced the warming trend in conceptual change. According to the CRKM model, the learner's level of engagement

* Corresponding author. Tel.: +1 215 204 6132; fax: +1 215 204 3592.
E-mail addresses: doug.lombardi@temple.edu, lombardi.doug@gmail.com (D. Lombardi).

0959-4752/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2013.03.001

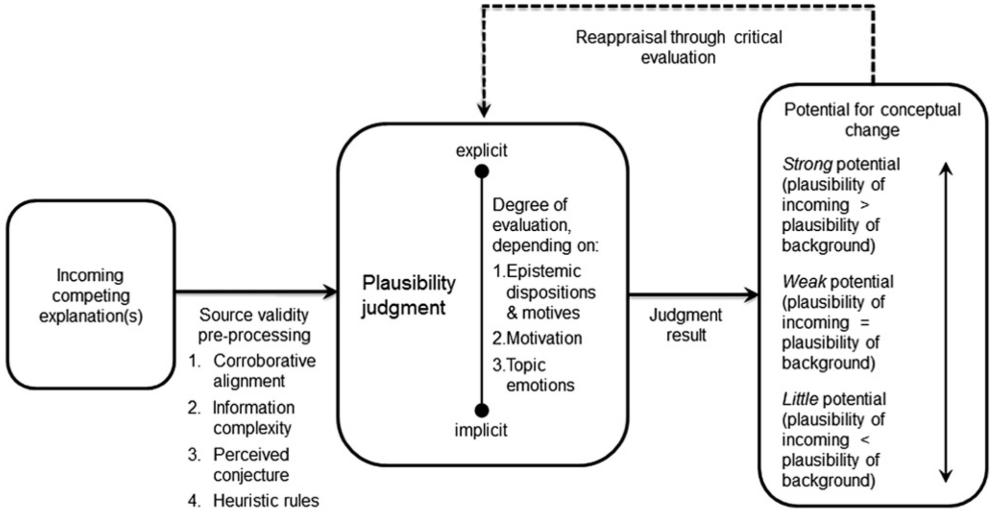

Fig. 1. A model of the role of plausibility judgments initiated by situations of competing explanations.

influences the likelihood of conceptual change. Level of engagement is determined by the nature of the interaction between the learner and message characteristics. Learner characteristics include depth and organization of background knowledge, motivational factors such as dissatisfaction with existing conceptions and personal relevance of the new information, and dispositions such as need for cognition, or the willingness to engage deeply with complex messages. Message characteristics refer to the learner's perceptions of the message, and specifically whether the learner finds the content comprehensible, coherent, *plausible*, and compelling. In the CRKM model, whether or not a message is perceived as plausible depends on the interaction between the learner's background knowledge (what and how much is known about the topic), the motivations to process or avoid processing the message, and characteristics of the message itself, such as comprehensibility.

Similar to the CRKM model, the plausibility model specifically incorporates motivation, emotion, and epistemic dispositions as important factors that influence the degree of cognitive evaluation in the formation of plausibility judgments. Lombardi et al. (in review) argue that plausibility judgments are often implicit and automatic cognitive processes; however, a plausibilistic comparison may be reappraised through explicit and effortful critical evaluation. Dole and Sinatra (1998) call this critical comparison *high metacognitive engagement*, which again reflects the theoretical consistency between the plausibility and CRKM models. However, this new plausibility model builds upon the CRKM model, specifically by providing more detail on how plausibility judgments are formed and can be reappraised through explicit cognitive processing. The plausibility model also moves beyond the CRKM model's notion of plausibility as a message characteristic. In the updated plausibility model, Lombardi et al. (in review) view plausibility as a tentative epistemic judgment that may be particularly influential on the conceptual change process in situations of competing explanations.

Research evidence in accord with theoretical models suggests that students need to explicitly reappraise their plausibility judgment in order to reconstruct their knowledge successfully on complex topics such as climate change, where initial considerations about scientific information may be based on thinking processes requiring low cognitive effort (Bråten, Strømsø, & Salmerón, 2011; Strømsø, Bråten, & Britt, 2010). Furthermore, the controversial nature of human-induced climate change may contribute to a perceived "plausibility gap" of individuals' conceptions relative to the scientific conception.

Based on these circumstances, the purpose of this study was to examine if critical evaluation of competing climate change models would help students re-evaluate their plausibility perceptions of human-induced climate change. This in turn could result in a greater likelihood of conceptual change about the topic. Next, we discuss alternative conceptions about climate change, the importance of plausibility judgments in conceptual change, and the connection between critical evaluation and plausibility reappraisal.

1.2. Alternative conceptions about the causes of climate change

Students and the general public alike hold several alternative conceptions about climate change (Leiserowitz & Smith, 2010). Cause-related alternative conceptions include (a) attributing global warming to increasing solar irradiance (i.e., the amount of solar energy received at the top of the Earth's atmosphere; see, for example, Boyes & Stanisstreet, 1993; Pruneau, Gravel, Courque, & Langis, 2003; Shepardson, Choi, Niyogi, & Charusombat, 2011), (b) stratospheric ozone depletion (i.e., the ozone hole) causing either increased amounts of energy to reach the Earth's surface or allowing more of Earth's energy to escape out to space (see for example, Boyes & Stanisstreet, 1993; Österlind, 2005), (c) a gas or dust layer at the top of Earth's atmosphere behaving similarly to a glass roof on a greenhouse (see, for example, Pruneau et al., 2003; Shepardson et al., 2011), and (d) some form of pollution (i.e., pollution other than greenhouse gas emissions) contributing to global warming (see, for example, Gowda, Fox, & Magelky, 1997; Papadimitriou, 2004). In the present study, we chose to address the alternative conception of increasing amounts of solar irradiation. This alternative conception is especially relevant to initial learning about climate change (i.e., at the middle school level) based on learning progressions implied by the recently developed framework for K–12 science education (National Research Council, 2012).

The sun is the predominant energy source for Earth's weather and climate. Recent paleoclimate studies have shown a strong association between solar activity and global temperatures over the past 11,000 years (Solanki, Usokin, Kromer, Schüssler, & Beer, 2004). However, "correlations between the Sun's behavior and the Earth's climate have completely failed since the 1970s" (Priest, Lockwood, Solanki, & Wolfendale, 2007, p. 3.7). Solar activity has been decreasing since that time, and in the absence of an enhanced greenhouse effect caused by human activities, this lessening solar irradiance should have resulted in slightly lower global temperatures (Lockwood, 2010).

Despite these recent scientific observations, the increased solar activity argument has been popular with those who are skeptical of human influences on climate (Cook, 2010). For example, a blog called the Dakota Voice misinterpreted a NASA study by claiming "we have still more evidence that any warming occurring on planet earth is coming from natural sources [i.e., the sun] and is cyclic in nature" (Ellis, 2009, p. 1). Educational researchers have also found that students hold alternative conceptions about the connection between climate change and solar irradiance. For example, Boyes and Stanisstreet (1993) found that 59% of secondary students ($N = 128$) incorrectly thought "the greenhouse effect is made worse because too many of the sun's rays get to the earth" (p. 538). Pruneau et al. (2003) surveyed 39 teenage students prior to instruction and found that a few believed that climate change is occurring because "the planet gets closer to the sun and gets warmer" (p. 437).

Both the blogger's and students' alternative conceptions may be related (at least in part) to the judgment that increased solar energy output is more *plausible* than the view that human activity can impact the earth's climate. We next turn to the connection between plausibility perceptions and climate change conceptions.

1.3. Empirical evidence of plausibility judgments in conceptual change

Conceptual change theorists have long hypothesized plausibility judgments as a critical component in knowledge reconstruction (Chinn & Brewer, 1993; Dole & Sinatra, 1998, Posner et al., 1982); however, plausibility has received little empirical attention. Treagust and Duit (2008) reported on a series of three studies supporting the idea that students must first comprehend the incoming message before they can make plausibility judgments. Interestingly, intentional learners (i.e., students with the goal of mastering the material) engage in deeper levels of processing (Sinatra & Taasoobshirazi, 2011), which could result in more reflection when making their plausibility judgments. With few exceptions, conceptual change researchers have not engaged in empirical investigations of plausibility perceptions until recently (see for example, Lombardi & Sinatra, 2012).

1.3.1. Plausibility judgments and reconstructing conceptions of climate change

Lombardi and Sinatra (2012) conducted a study with 83 undergraduate students and found that plausibility perceptions about human-induced climate change accounted for statistically significant increases in knowledge about weather and climate distinctions (a common source of confusion in climate change) over semester-long instruction, above and beyond their existing background knowledge. Lombardi and Sinatra also found that plausibility perceptions did not significantly change during instruction, even though one of the courses involved in the study focused on climate science for the entire semester. Lombardi and Sinatra hypothesized that students' plausibility judgments did not shift over the course of the semester because they were not given the opportunity to weigh "the plausibility of geoscientists' claims with alternative claims" (Lombardi & Sinatra, 2012, p. 212). This suggestion helped inform the preliminary development of a model of plausibility judgments (see Fig. 1; Lombardi et al., in review).

The model shown in Fig. 1 highlights the potential importance of the "plausibility appraisal through critical evaluation" as a feedback loop. In the case of science learning, critical evaluation involves judgments about the relationship between evidence and alternative explanations of a particular phenomenon (McNeill, Lizotte, Krajcik, & Marx, 2006). Critical evaluation is akin to a problem solving process where individuals engage in metacognitive reflection that may facilitate revision of existing plausibility judgments (Pintrich et al., 1993). Critical evaluation would potentially promote systematic reappraisal of plausibility "allowing for a revised appraisal of the initial data" (Rescher, 1976, p. 119) that could—in turn—result in conceptual change. Lombardi and Sinatra (2012) speculated that if the students had engaged in critical evaluation, their plausibility perceptions about human-induced climate change may have increased, with a subsequently potential greater reduction in their alternative conceptions about the distinctions between weather and climate. A primary motivation for the present study is to provide a direct test of the Lombardi et al. (in review) model, and specifically to examine if critical evaluation results in plausibility reappraisal and conceptual change.

1.3.2. Factors influencing the degree of evaluation in plausibility judgments

Lombardi and Sinatra (2013) have also examined some factors relating to teachers' initial plausibility judgments and found evidence that the degree of evaluation in making plausibility judgments is influenced by both topic emotions—emotions that relate specifically to the topic of instruction—and epistemic motives—specifically, the need for closure, which represents individuals' "motivation with respect to information processing and judgment" (Webster & Kruglanski, 1994, p. 1049). With regard to topic emotions, greater anger about climate change predicted lower plausibility perceptions and greater hopelessness about climate change predicted greater plausibility perceptions. Lombardi and Sinatra also found that anger about teaching climate change and decisiveness (a need for closure subcomponent reflecting a desire for definitive answers) were significant predictors, with both greater anger and greater decisiveness resulting in lower plausibility perceptions of human-induced climate change. Lombardi and Sinatra's (2013) study was limited because they did not measure conceptual knowledge postinstruction, and therefore, they were unable to determine if any conceptual change occurred. However, we can speculate that if these participants had engaged in critical evaluation, they may have reappraised their plausibility judgment and potentially had greater plausibility perceptions about human-induced climate change.

1.4. Reappraising plausibility through critical evaluation

Halpern (2007) lists many attributes of critical thinking and specifically states that critical thinking "involves evaluating the thinking process—the reasoning that went into the conclusion we've arrived at or the kinds of factors considered in making a decision" (p. 5). However to employ critical evaluation, an individual must examine the connection between evidence and explanation, as well as connections between the same evidence and alternative explanations (McNeill et al., 2006).

1.4.1. The need for explicit evaluation

Students may be naturally curious about scientific topics, but are not necessarily evaluative as they consider hypotheses and theories. Chinn and Buckland (2012) state that some students adopting a creationist perspective on biological evolution may engage in non-collaborative argumentation tactics that bias evidence. Such a stance may prevent learning the central tenants of biological evolution (Chinn & Buckland, 2012). To overcome this bias and promote deeper learning of evolution, Chinn and Buckland argue that students should gain a coordinated understanding of both the theory's conceptions and scientists' epistemic practices. Scientific judgments about the strength of the theory of biological evolution are based on a large body of evidence. Furthermore, these judgments have emerged from an environment of argumentation that has considered alternative explanations (e.g., intelligent design).

Critical evaluation involves understanding how evidence relates to an idea (e.g., an argument, a scientific model) and its alternatives (e.g., a counterargument, a contrary hypothesis) (McNeill et al., 2006). Through critical evaluation, an individual seeks to weigh the strengths and weaknesses in the connection between the evidence and the ideas. Mere critique is not sufficient. For example, people can exhibit a disconfirmation bias, "where when faced with evidence contrary to their beliefs, people try to undermine [this incoming] evidence" (Edwards & Smith, 1996, p. 6). The purpose of this undermining memory search is to "retrieve material [e.g., stored beliefs] for use in refuting the position advocated" (Edwards & Smith, 1996, p. 18). However, this disconfirmation bias is not necessarily evaluative because less cognitive processing is involved when individuals agree with a particular position. Therefore, critical evaluation must try to find fault with both the existing idea and the alternative, gauged on the level of support provided by evidence. In this way, critical evaluation embraces the scientific standard of falsifiability (Popper, 1963).

1.4.2. Instruction promoting critical evaluation

Students' classroom use of critical evaluation should mimic that used by scientific experts (Duschl, Schweingruber, & Shouse, 2007). By publishing their work in research journals and participating in symposia, panels, and presentations, the scientific community engages in collaborative argumentation, defined by Nussbaum (2008) as a constructive and social process where individuals work together to compare, critique, and revise conceptions. However, students may not naturally be critically reflective when engaging in collaborative argument, and therefore, "students need tools to evaluate arguments" (Nussbaum & Edwards, 2011, p. 447), such as use of critical questions (Nussbaum & Edwards, 2011), knowledge of epistemic criteria and disciplinary norms (Duschl, 2008), knowledge of content and arguments on both sides of an issue (Kardash & Scholes, 1996), appreciation of the role of criticism (Szu & Osborne, 2012), and graphic organizers and other supports to help distinguish, coordinate, and evaluate theory and evidence (Cavagnetto & Hand, 2012).

Science topics, however, may involve additional complexities for learning. Specifically, students encountering complex science topics—such as climate change—may possess existing mental representations that conflict with scientific understanding, and often, these naïve understandings seem more plausible to them than the correct conception. For students to be able to critically judge plausibility when comparing competing models (naïve versus scientific), they need to weigh evidentiary data (Chin & Osborne, 2010).

Chinn and Buckland (2012) report on the recent use of an instructional scaffold, called the model-evidence link (MEL) diagram, which assists students in making arguments based on the relative weighting of evidence that support an explanatory model and an alternative. In a year-long study involving middle school life science students, Chinn, Duschl, Golan Duncan, Buckland, and Pluta (2008) found that the treatment group (students using MEL diagrams) made "substantially greater advances in their ability to effectively coordinate models and evidence" than the comparison group (students who did not use the modeling scaffold) (p. 2). Each MEL diagram used in the study featured two models of ulcer inducement (e.g., the naïve stress model versus the scientific bacteria model). Students would gather evidentiary data during the instructional activities (e.g., reading a passage about a scientific experiment) and collaborate on constructing MEL diagrams. Participants in the control group completed the same argumentation activities, but did not use the MEL diagrams. Many of these units involved topics for which students typically have robust alternative conceptions (e.g., photosynthesis, cellular respiration, and mitosis). Pre and post testing for each unit demonstrated that treatment group participants experienced a greater degree of conceptual change than control group participants (Chinn & Buckland, 2012), thus supporting the idea that the MEL is an effective instructional scaffold. In the present study, we used the MEL to promote critical evaluation, influence plausibility reappraisal, and facilitate conceptual change, per the model developed by Lombardi et al. (in review) (Fig. 1).

1.5. The present study

Based on the research highlighted above, we hypothesized that students would increase their plausibility perceptions of human-induced climate change when engaging in critical evaluation of competing climate change models. In addition, critical evaluation and reappraised plausibility judgments would lead students to restructure their knowledge about human-induced climate change.

1.5.1. Research question

We developed the following research question for the present study: When engaged in instruction designed to promote critical evaluation, do students' (a) plausibility perceptions of competing climate change models and (b) knowledge about this topic change? The instruction focused on students' evaluation of two competing models: human-induced climate change (i.e., the scientifically accurate model; Intergovernmental Panel on Climate Change, 2007) and increasing solar energy (i.e., a popular skeptic model; Cook, 2010; Ellis, 2009).

1.5.2. Hypotheses

We hypothesized that students who used MEL diagrams to critically evaluate evidence for each model would rank the plausibility of the scientifically accurate conception higher than the alternative conception and would reappraise their preinstruction plausibility (Hypothesis 1a). This hypothesis concerns the relationship between critical evaluation and plausibility reappraisal (i.e., changes in plausibility perceptions), and is consistent with Lombardi et al.'s (in review) model of plausibility judgments, which suggests that critical comparison of alternative explanations can move plausibility judgments from implicit processing to explicit reflection. This hypothesis is also consistent with Dole and Sinatra's (1998) CRKM model, which suggests that explicit opportunities to "think deeply about the arguments and counterarguments" (p. 121) leads to higher engagement and greater likelihood of conceptual change. We also hypothesized that using MEL diagrams as an instructional tool would result in students' restructuring their knowledge about human-induced climate change and that such instruction would result in strong and enduring conceptual change (Hypothesis 1b). We based this hypothesis on the theoretical conceptual change model of Dole and Sinatra (1998) and empirical research by Lombardi and Sinatra (2012) linking plausibility judgments to change, as well as the recently developed model of plausibility judgments implicating plausibility reappraisal in facilitating change (see Fig. 1; Lombardi et al., in review).

2. Methods

2.1. Participants and setting

Middle school students from a large, urban, ethnically diverse district in the Southwestern USA served as participants. The school district involved in this study teaches about climate during grade 7, when all students are required to take an Earth science class. Study participants were drawn from an entire middle school's grade 7. These participants were enrolled in Earth science and were taught by one of four science teachers. At the time of the study, 429 students were enrolled in Earth science and we invited all to participate in the study. About 63% ($N = 269$) of the students provided both parental consent and self-assent. Most of the students provided self-assent because they signed these forms in class. However with the exception of one student (who returned a form with parental non-consent clearly indicated), other non-consents resulted from failure to return a signed consent form from home. The teachers indicated that such behavior (difficulty in returning material from outside the classroom, such as homework) is typical. Despite encouragement from the teachers to return consent forms, many students may not have done so because they knew there would be no negative consequence (i.e., the students knew that although the activity would be conducted for all students because of its alignment with district curriculum and instructional goals, there would no penalty for not returning forms and lack of consent).

Just under two-thirds ($N = 169$) fully participated in the preinstruction, quasi-experimental, and postinstruction phases. Of the 169 students who participated, 108 (64%) were Hispanic, 29 (17%) were White, 19 (11%) were African American, and 13 (8%) were Asian/Pacific Islander. Eighty-seven participants (51%) were male. Eighteen (11%) of the participants had individualized education plans, 36 (21%) had limited proficiency in the English language, and 79 (47%) were eligible for free or reduced-cost lunch. Demographic composition was similar between the treatment group (64% Hispanic; 16% White; 15% African American; 5% Asian/Pacific Islander; 52% male; 9% individualized education plans; 21% limited proficiency in the English language; and 47% eligible for free or reduced-cost lunch) and comparison group (64% Hispanic; 18% White; 7% African American; 11% Asian/Pacific Islander; 51% male; 12% individualized education plans; 22% limited proficiency in the English language; and 47% eligible for free or reduced-cost lunch).

2.2. Design and materials

The preinstruction, quasi-experimental, and postinstruction phases were conducted toward the end of the school year's first quarter. At this time, the grade 7 students were completing an introductory unit on the nature of Earth science. The instructional activities occurred over two class periods (about 90 min of instructional time total). Fourteen total classes were involved in the study (three different teachers were instructors for four classes each and one teacher was the instructor for two classes). We randomly assigned half of the classes to the treatment condition (i.e., using an instructional activity promoting critical evaluation of two competing climate change models) and the other half of the classes to the comparison condition (i.e., using regular curriculum materials that discuss climate change). The second part of the study—the delayed postinstruction phase—occurred six months later.

2.2.1. Perceptions of model plausibility

Two items measured the plausibility evaluation of Model A (human-induced climate change) and Model B (solar irradiance causing climate change). These items asked the participants to rate the plausibility of each model using a 1–10 scale, where 1 = greatly implausible or even impossible and 10 = highly plausible. In measuring plausibility, we directly examined perceptions of potential truthfulness of two claims explaining the cause of current climate change (i.e., Model A and Model B). We constructed the model plausibility items similar to our more general plausibility perceptions measure (PPM; Lombardi & Sinatra, 2012). The PPM is an instrument that broadly measures plausibility of several scientific statements about climate change (e.g., claims about observational evidence, causes of current climate change and future impacts of climate change). However, in the present study, we were interested in focusing directly on the plausibility perceptions that participants had about the two alternative models, which is a straightforward judgment about the claim made in a particular model statement about the causes of current climate change (e.g., Model A claims that current climate change is caused by increasing amounts of gases released by human activities). Plausibility judgments about other statements (e.g., scientific predictions about future impacts of climate change) are not directly related to the models in the present study, and therefore, additional plausibility items (e.g., such as those found in the PPM) would not be relevant to the claims made by these models.

We gauged reliability of model plausibility perceptions using the Spearman Brown coefficient, which is the most appropriate statistic for two-item measures (Eisinga, Grotenhuis, & Pelzer, 2012). The Spearman Brown coefficient was equal to .696, which is right at the threshold that is commonly considered acceptable (George & Mallery, 2009). Participants completed these model ratings during the preinstruction phases after reading a one-page text introduction of the two alternative models and the notion of plausibility. Model ratings were also completed at postinstruction.

2.2.2. Knowledge of human-induced climate change

We developed a 27-item instrument to measure participants' knowledge of human-induced climate change (HICCK) just prior to, immediately after instruction, and six months after instruction. We created this instrument to measure conceptions about the current scientific consensus on human-induced climate change based on a recent study that surveyed American citizens on their understanding of scientific phenomena related to global warming (Leiserowitz & Smith, 2010), the latest summative report produced by a United Nations' expert panel (Intergovernmental Panel on Climate Change, 2007), and common alternative conceptions about human-induced climate change (Choi, Niyogi, Shepardson, & Charusombat, 2010). The participants rated each item on a 5-point Likert scale gauging the level of agreement that they thought *climate scientists* would indicate for each statement, ranging from 1 = strongly disagree to 5 = strongly agree, for example, "current climate change is caused by an increase in the Sun's energy" (see Appendix 1 for full survey). We should note that using a scale measuring the level of agreement with climate scientists allowed us to gauge understanding of scientific explanation, not acceptance of or attitudes toward scientific explanations.

As DeVellis (2003) recommends, HICCK items were strongly worded, unambiguous declarative statements without jargon. The Flesch–Kincaid formula indicates that readability of HICCK items is slightly below the grade 7 level, on average. Five of the HICCK items directly address alternative conceptions about the causes of climate change, as summarized by Choi et al. (2010). Overall reliability of the HICCK for the preinstruction, postinstruction and delayed postinstruction administrations was at the acceptable threshold (Cronbach's $\alpha = .687$; George & Mallery, 2009).

2.2.3. Instructional scaffold

The treatment group used the model-evidence link (MEL) diagram activity to promote critical evaluation and potential reappraisal of plausibility judgments about human-induced climate change. On a MEL provided to each student, participants drew different types of arrows linking evidentiary data to the two alternative models of climate change (Model A: human-induced and Model B: solar irradiance). Participants drew arrows in different shapes to indicate the relative weight of the evidence. Straight arrows indicated that evidence supports the model; squiggly arrows indicated that evidence strongly supports the model; straight arrows with an “X” through the middle indicated the evidence contradicts the model; and dashed arrows indicated the evidence has nothing to do with the model (see Fig. 2 for a student-completed MEL).

Prior to the present study, we conducted two pilot studies using different versions of the MEL. Grade 7 students from another middle school participated in the pilot studies. The school used in the pilot study had similar demographics to that of the present study. In the first pilot study, we tested a MEL with six evidence statements (i.e., two more than we used in the present study). Nonparametric statistical tests showed no significant differences, which was likely due to problems with the preliminary version of the MEL used. Specifically, a MEL with six evidences may have resulted in cognitive overload as students attempted to engage in critical evaluation. For the second pilot study, we developed a MEL with only four statements that cover essential evidence related to the two alternative models in the MEL (Intergovernmental Panel on Climate Change, 2007). Results of this test showed that perceived model ratings changed significantly from pre to postinstruction, and based on this significant gain, the adjustments we made in the MEL may have reduced cognitive overload and allowed for clearer connections between evidentiary data and the models. This four-evidence version of the MEL was used in the present study.

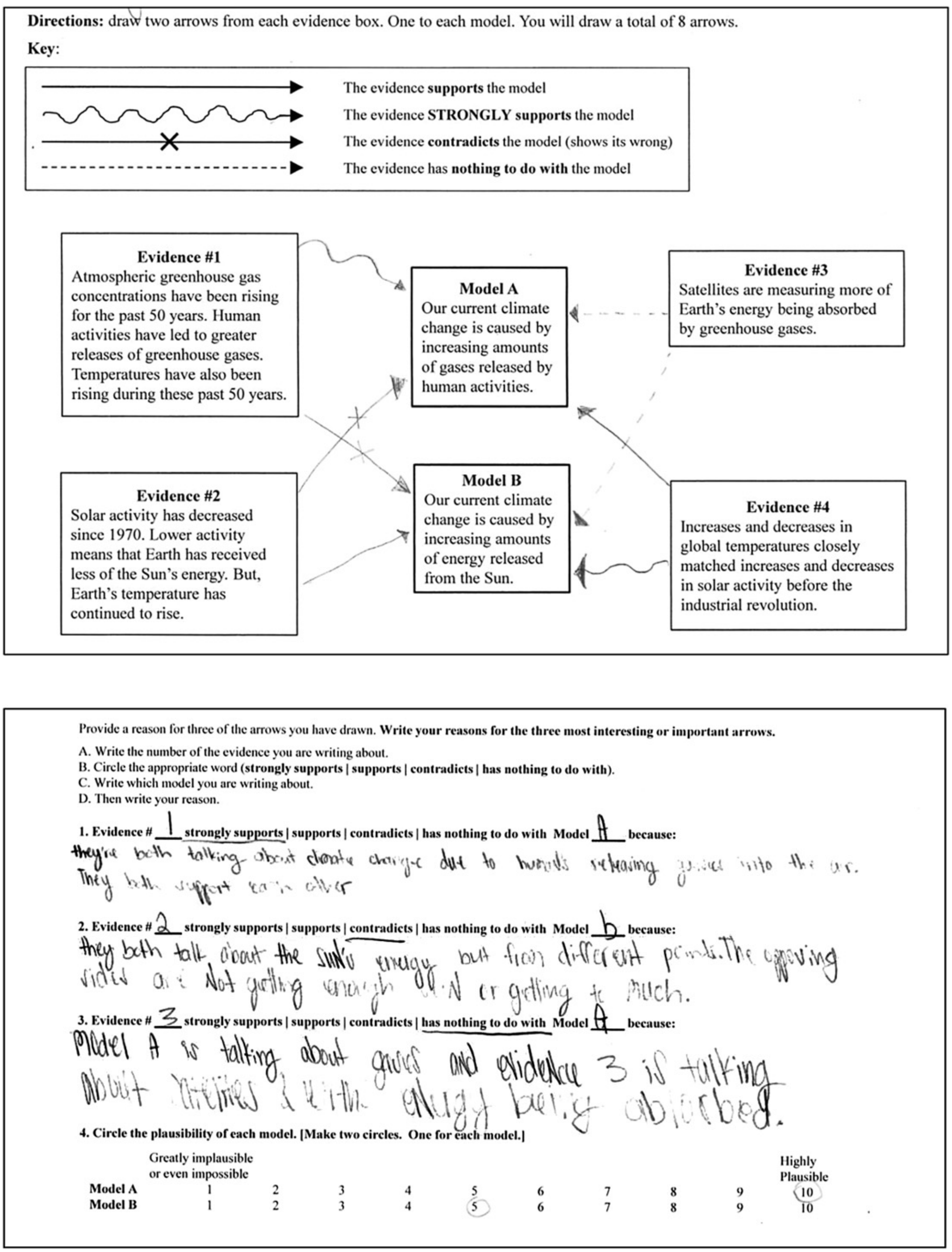

Fig. 2. Student example of a model-evidence link (MEL) diagram.

2.3. Procedures

Prior to the instructional activity, participants completed the human-induced climate change knowledge (HICCK) instrument and climate change model ratings of plausibility. Before completing the pre ratings of model plausibility, participants read a short introduction to the two models and a statement defining plausibility. This short introduction exposed both the treatment and comparison groups to the two models and the notion of plausibility. Each teacher conducted a short discussion to help the students clarify any misunderstandings about the models and plausibility. Both treatment and comparison groups participated in this short discussion for about the same amount of time.

Participants in the treatment group engaged in a MEL diagram activity that was taught by their regular classroom teacher. The first part of the activity was titled "How do scientists change their plausibility judgments?" Our intent was to help students understand how scientists weigh connections between evidence and scientific ideas (e.g., scientific models). Specifically, this part asked students to rank the importance of the following four types of evidence connections in changing plausibility judgments:

1. The evidence supports an idea.
2. The evidence strongly supports an idea.
3. The evidence contradicts (opposes) an idea.
4. The evidence has nothing to do with the idea.

Note that these statements correspond to the four types of arrows that the participants used when they developed their MELs (see Fig. 2).

After making their initial rankings, the treatment participants read a short paragraph discussing falsifiability, and specifically, how evidence that contradicts an idea has a large influence on how scientific knowledge changes. Participants then re-ranked the four types of evidence. After re-ranking, teachers conducted a short discussion with the class on their rankings and directly reinforced that contradictory evidence generally has the greatest weight in changing scientists' plausibility judgments.

In the next part of the MEL activity, the instructor had the treatment participants individually read short expository texts discussing each piece of evidence (i.e., one page of text for each evidence). These pages also included graphs and figures. The instructors asked the students if they had any questions about the evidence texts, figures, and graphs to clear up any confusion or misunderstandings. Treatment group participants evaluated the four evidentiary statements and linked them to each model using different arrows for the weighting scheme. After completing their diagrams, treatment participants individually completed a short explanatory task, which allowed students to reflect on the arrows they drew on the MEL, and then participants rated each model's plausibility individually (i.e., the same as they did during preinstruction). These explanatory tasks were similar in structure to the investigation question tasks used in the comparison group activity.

Comparison group participants used instructional materials from *Integrating Earth Systems* (IES) Weather and Climate module (Smith, Southard, & Mably, 2002). The IES instructional materials were developed through funding provided by the National Science Foundation and the American Geological Institute. According to the material developers, an independent evaluator found that use of the IES materials "led to significant gains in student understanding of fundamental Earth science concepts" (Smith et al., 2002). Engaging students in scientific inquiry was one of the principle design considerations of IES and the curriculum specifically endeavors for students to build evidence to make explanations about Earth systems that are consistent with scientific understanding. IES represents instructional materials that are consistent with science education reform efforts of the middle and late 1990s, and early 2000s (National Research Council, 1996). Specifically, "IES emphasizes the importance of learning science through inquiry" (Penuel & Gallagher, 2009, p. 472) rather than through didactic learning typical of traditional textbooks.

The IES weather and climate module has eight investigations. We specifically used Investigation 8, titled "How is Global Climate Changing," for the comparison activity. In this activity, comparison group participants were initially asked the following guiding questions: "Do you think the world's climate is changing? If so, what will happen in the future? What will the climate be like for you, your children, and your grandchildren?" Comparison group participants then read about evidence related to past and current climate change, and then made predictions about future climate change as a collaborative group.

The investigation was adapted so that comparison group participants would read and use the same four pieces of evidence used in treatment activity. By using the same text in the treatment and comparison activities, we made the instructional experience as similar as possible, (i.e., both groups had the same information source). These evidence texts were used to answer questions throughout the investigation so that comparison group participants could evaluate these evidences. For example, two questions ask the participants to consider "What parts of the four evidences support your final prediction? What parts of the four evidences do not support your final prediction?" Whereas such questions are evaluative, Investigation 8 did not ask the participants to weigh evidence between two competing models. This is the critical difference between the comparison task and the treatment task (i.e., the MEL diagram activity). The time needed by comparison group participants to complete Investigation 8 was two class periods (i.e., same amount of time spent by the treatment group participants on the MEL diagram activity). Furthermore, all teachers interacted with both comparison and treatment participants in the same manner as the participants completed their activities (i.e., answering clarifying questions, encouraging use of the evidence texts, ensuring students focused on the instructional tasks, etc.)

At the end of the learning activities, treatment and comparison group participants completed the HICCK for a second time. Treatment group participants also completed the two items measuring comparative plausibility at the end of the MEL diagram activity, whereas comparison group participants completed these two items using the same instrument as in preinstruction. For both the treatment and instruction groups, completion of measures at preinstruction was completed over two 45 min class periods. The instructional activities for both the treatment and comparison groups were also completed in two 45 min class periods; likewise completion of measures at postinstruction also took two 45 min class periods for both groups. Six months later, participants in both the treatment and comparison groups completed the HICCK for the third and final time. Teachers administered this final HICCK just prior to conducting a lengthy unit specifically focused on weather and climate. Prior to this time, the only instruction on climate was associated with activities that students completed during the quasi-experimental phase. Only three teachers were able to administer the HICCK at delayed postinstruction due to scheduling conflicts, which lowered the number of participants ($N = 96$) for this final phase.

3. Results

Table 1 shows the means and standard deviations for perceptions of model plausibility (Mplaus) and knowledge of human-induced climate change (HICCK) by time period (pre and post-instruction), as well as group (treatment and comparison).

Table 1
Means and standard deviations (in parentheses) for the study variables at pre and postinstruction ($N_{treatment} = 86$, $N_{comparison} = 83$, $N_{total} = 169$).

Variable	Group	Preinstruction	Postinstruction
Mplaus	Treatment	−0.30 (3.99)	1.60 (2.82)
	Comparison	0.04 (3.85)	−0.19 (3.61)
	Overall	−0.14 (3.91)	0.72 (3.34)
HICCK	Treatment	92.3 (8.65)	95.5 (10.1)
	Comparison	91.5 (8.27)	90.7 (8.84)
	Overall	91.9 (8.45)	93.2 (9.77)

Note. The possible score ranges were: (a) perceptions of model plausibility (Mplaus) = −9 to +9; and (b) knowledge of human-induced climate change (HICCK) = 34–170.

3.1. Assumptions testing

We performed a repeated measures multivariate analysis of variance (MANOVA) to assess changes pre to postinstruction, with group (treatment and comparison) as the between-subjects variable, time (pre and postinstruction) as the within-subjects variable, and Mplaus and HICCK as the dependent variables. We screened the results and found that our analysis met the normality, linearity, and homogeneity of the variance-covariance matrices assumptions inherent in MANOVA designs.

Individual participants were nested in classrooms, creating possible statistical dependencies among students within these classrooms. We therefore calculated the intra-class correlations (ICC) to ascertain levels of statistical dependence among the observations. A common way to determine ICC is to calculate the proportion of variance explained by group membership, use the calculated ICC value to calculate the test statistic (z), and then determine the p-value from the test statistic (Kashy & Kenny, 2000). Table 2 shows the ICC values that we calculated for the two dependent variables (all using postinstruction scores), with associated one-way ANOVA parameters (with teacher as the independent variable category) and calculated z- and p-values. Because all dependent variables had no significant ICC values and ICC values were less than .2, the participants in a particular teacher cluster did not have significantly greater similarity than the overall similarities in the treatment and comparison groups (Cress, 2008; Snijders & Bosker, 1999). This indicates that differences between the treatment and comparison groups were not likely due to potential classroom effects and that our estimated standard errors are valid.

3.2. Pre to post differences

As a reminder, we used a repeated measures MANOVA with model plausibility and knowledge of human-induced climate change as the dependent variables. The repeated measures MANOVA revealed a significant interaction between group and time for the combined scores of perceptions of model plausibility and knowledge of human-induced climate change, with $F(2,166) = 9.01$, $p < .001$, $\eta^2_p = .098$ (a medium effect size, Tabachnick & Fidell, 2007).

Follow-up univariate analyses of variance indicated that interactions between time and group were significant for perceptions of model plausibility, $F(1,167) = 10.89$, $p = .001$, $\eta^2 = .061$ (medium effect size), and knowledge of human-induced climate change $F(1,167) = 9.26$, $p = .003$, $\eta^2 = .053$ (small effect size). These analyses were run concurrently, and to properly account for familywise error, we used an adjusted critical value ($\alpha = .025$) as a conservative gauge of significance.

We conducted a simple effects analysis to determine the exact nature of the group differences at both preinstruction and postinstruction. There were no significant differences preinstruction between the treatment and comparison groups in either variable, with all p-values > .53. However at postinstruction, treatment group scores were significantly greater than the comparison group scores in both variables, with $F(1,167) = 13.09$, $p < .001$, $\eta^2 = .073$ (medium effect size) for participants' perceptions of model plausibility, and $F(1,167) = 10.67$, $p = .001$, $\eta^2 = .060$ (medium effect size) in scores of knowledge of human-induced climate change.

The simple effects analysis also showed that there were no significant differences in any of the comparison group scores, pre to postinstruction, with all p-values > .43. However when comparing both variables from preinstruction to postinstruction, the treatment group had statistically greater scores in perceptions of model plausibility and knowledge of human-induced climate change (all p-values < .001; see Table 1 for means and standard deviations).

3.3. Delayed post differences

In order to examine sustained knowledge change, we conducted a repeated measures analysis of variance (ANOVA) with group (treatment and comparison) as the between-subjects variable, time (pre, post, and delayed postinstruction) as the within-subjects variable, and HICCK scores as our dependent measure. Three of the four teachers involved in the study were able to re-measure participants' knowledge of human-induced climate change about six months after instruction. Demographic composition of this reduced number of participants ($N = 96$; 67% Hispanic; 16% White; 10% African American; 7% Asian/Pacific Islander; 46% male; 13% individualized education plans; 21% limited proficiency in the English language; and 44% eligible for free or reduced-cost lunch) was similar to the larger number of participants discussed above.

Table 3 shows the means and standard deviations for both treatment ($N = 48$) and comparison ($N = 48$) groups by time period (pre, post, and delayed postinstruction) for participants taught by these three teachers. Fig. 3 graphically summarizes the pre to post to delayed postinstruction changes in HICCK scores.

To examine changes in HICCK scores, we conducted a repeated measures analysis of variance (ANOVA) with group (treatment and comparison) as the between-subjects variable, time (pre, post, and delayed postinstruction) as the within-subjects variable. The results show a non-significant interaction between group and time in HICCK scores, with $F(2,183) = 2.65$, $p = .074$. However with a relatively low p-value, follow-up simple effects can still reveal statistically significant differences (Clark-Carter, 2010), which was the case with the present study. The simple effects analysis on the repeated measures ANOVA results revealed a statistically significant increase from pre to postinstruction in treatment group participants' scores, with $F(2,93) = 3.24$, $p = .044$, $\eta^2 = .065$ (medium

Table 2
Intra-class correlation coefficient (ICC) values for the study variables clustered by teacher, with $N_{teacher} = 4$, and m (mean cluster size) = 21.13.

Variable	MS_B	MS_W	ICC	z	p-Value
Mplaus	5.218	8.035	0.0169	0.0719	0.943
HICCK	285.827	95.317	0.0864	0.3689	0.714

Note. Mplaus = perceptions of model plausibility; HICCK = knowledge of human-induced climate change; MS_B = mean sum of squares between clusters; and MS_W = mean sum of squares within.

Table 3
Means and standard deviations (in parentheses) for the human-induced climate change knowledge scores (HICCK) at pre, post, and delayed postinstruction ($N_{treatment} = 48$, $N_{comparison} = 48$, $N_{total} = 96$).

Group	Preinstruction	Postinstruction	Delayed postinstruction
Treatment	89.9 (8.97)	92.8 (9.67)	92.6 (8.61)
Comparison	88.7 (7.07)	88.2 (6.74)	88.0 (7.78)
Total	89.3 (8.06)	90.5 (8.61)	90.3 (8.48)

Note. The possible score ranges were 34–170.

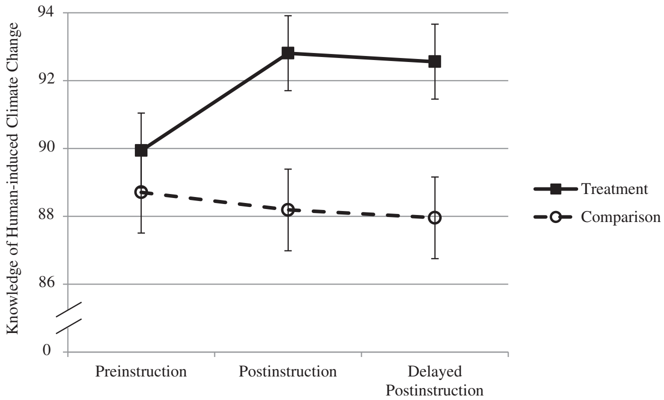

Fig. 3. Pre to post to delayed postinstruction human-induced climate change (HICCK) scores for the treatment and comparison groups, with bars showing standard errors. Note that the number of participants ($N = 96$) was reduced from comparisons showing pre and postinstruction changes only because one of the teachers in the study could not administer the instrument at the delayed post time period.

effect size). Furthermore, there was no significant change in treatment group participants' scores from post to delayed postinstruction ($p = 1.0$), indicating that the treatment group retained their statistically significant increase in knowledge six months after instruction. Treatment group HICCK scores were also significantly greater than comparison group scores at both post ($p = .008$) and delayed postinstruction ($p = .007$). However, there was no significant difference between the groups at preinstruction, $p = .46$, and comparison group scores remained unchanged at this relatively low level (i.e., no statistically significant differences) from pre to post to delayed postinstruction, with $p = .85$.

3.4. Specific indicators of conceptual change

We used the knowledge of human-induced climate change (HICCK) instrument to measure participants' understanding about the causes of climate change, and as the results showed, the treatment group participants experienced a significant change in understanding from pre to postinstruction and sustained this understanding up to six months later. Six of the items on the HICCK directly examined knowledge about the causes of current climate change and the potential for a conceptual shift in understanding about these causes. One of these items reflected the scientific model that humans are the current cause of climate change (i.e., the correct conception). The other five items looked at alternative conceptions about the causes of climate change; i.e., current climate change is caused by (a) an increase in the Sun's energy, (b) the ozone hole, (c) changes in Earth's orbit around the Sun, (d) volcanic eruptions, and (e) increasing dust in the atmosphere (Choi et al., 2010).

Fig. 4 shows how participants changed their conceptions on these six items, pre to postinstruction. Change is shown for the treatment and comparison groups for each item and is expressed as the mean change score (mean postinstruction score minus mean preinstruction score). Positive scores for the first item indicated a change toward understanding that is consistent with the scientific conception (current climate change is caused by humans). For the remaining five items, negative change scores indicated an overall reduction in misunderstanding about the causes of climate change (i.e., a lessening of alternative conceptions).

The treatment group had a positive mean change score on the scientifically correct item—indicating acquisition or strengthening of the correct conception—and negative mean change scores on the remaining five items—indicating a reduction in alternative

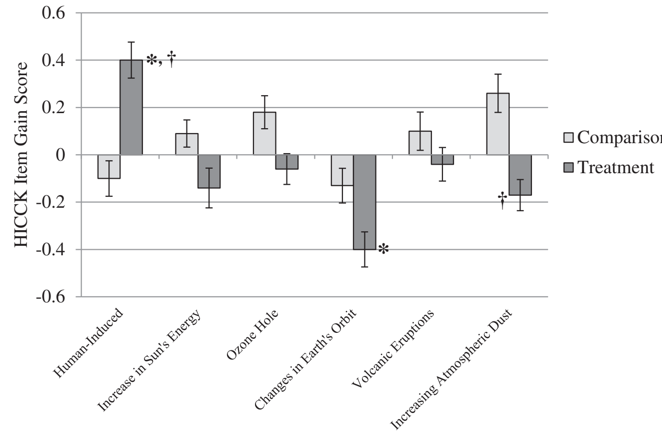

Fig. 4. Change scores (postinstruction – preinstruction) on six items from the knowledge of human-induced climate change (HICCK) instrument relating to causes of current climate change. For the first item (human-induced), positive scores indicate a change toward the scientifically accurate conception. For the other items, negative scores indicate a reduction in alternative conceptions. The two treatment group gain scores with an asterisk (*) are statistically significant change scores (assuming $\alpha = .05$), with both p-values = .007. The daggers (†) indicate statistically significant differences (assuming $\alpha = .05$) between the treatment and comparison groups, with both p-values < .03. Bars show standard errors.

conceptions . However these changes were only significant on two items as measured by paired t-tests. On the item measuring understanding of the scientific model, postinstruction scores ($M = 3.70$, SD = .94) were significantly greater than preinstruction scores ($M = 3.28$, SD = 1.24), $t(85) = -2.75$, $p = .007$, Cohen's $d = .30$. This result shows that treatment group participants' experienced significant conceptual change toward a view that is consistent with that of climate scientists. Treatment group participants also significantly reduced their alternative conception that climate change is caused by variations in Earth's orbit around the Sun, $t(85) = -2.74$, $p = .007$, Cohen's $d = .29$. None of the other postinstruction changes made by treatment group participants were significantly different from preinstruction, with all p-values > .19.

Comparison group participants had no significant differences in any of the items pre to postinstruction. Furthermore, independent-measures t-tests revealed a significant difference on the item measuring understanding of the scientific model between treatment group change scores ($M = 0.42$, SD = 1.41) and comparison group change scores ($M = -0.10$, SD = 1.37), $t(167) = -2.41$, $p = .017$, Cohen's $d = .37$. There was also a significant difference in change scores on the alternative conception that increasing atmospheric dust is causing current climate change between the treatment group ($M = 0.17$, SD = 1.22) and the comparison group ($M = -0.29$, SD = 1.47), $t(167) = -2.24$, $p = .027$, Cohen's $d = .34$.

3.5. Summary of results

Overall, the results showed that treatment group participants experienced significant changes in their perceptions of model plausibility and knowledge of human-induced climate change after experiencing instruction that promoted critical evaluation. These changes represented medium effect sizes and show that participants moved toward greater plausibility perceptions and knowledge of the scientific model of human-induced climate change. Additionally, knowledge change was sustained up to six months after instruction. The comparison group, which experienced the regular curriculum, did not show any significant changes in these variables.

4. Discussion

Our findings showed support for our hypotheses (Hypotheses 1a and 1b) that instruction promoting critical evaluation and plausibility reappraisal may facilitate sustained conceptual change. The instructional scaffold used in this study (the MEL diagram), where students weigh the links between scientific evidences and two alternative models of climate change, shows promise for providing students the instructional scaffold needed to critically evaluate evidence, shift their plausibility judgments, and change their conceptions. This change was demonstrated to maintain even six months after instruction.

This study expands our understanding of how plausibility perceptions may influence conceptual change. Several conceptual change models consider plausibility to be a key construct contributing to whether or not individuals change their knowledge (Chinn & Brewer, 1993; Dole & Sinatra, 1998). However, few studies have examined plausibility's role in conceptual change empirically. The present study is unique in that it is the first study to our knowledge to directly examine the impact of an intervention designed to shift plausibility perceptions on conceptual change outcomes. The results suggest that, as several models have predicted, plausibility may be a key component of conceptual change, particularly when the concept is one where there is a significant "plausibility gap" between scientific and lay judgments.

The study also has implications for the development of epistemic cognition. Our results suggest that reconstructing knowledge about human-induced climate change is neither a simple matter of debunking non-scientific positions nor learning about the several lines of evidence that support the scientific model. Rather, moving toward the scientifically accepted conception that humans are altering Earth's climate was facilitated by connecting evidences to alternative models and evaluating the strength of these connections with respect to each alternative. These are epistemic practices engaged in by scientists in the course of their work. The process of critical evaluation and plausibility reappraisal may help promote "epistemic conceptual change," as called for by Sinatra and Chinn (2012)—as well as the development of scientific habits of mind needed for an informed citizenry—if incorporated into the science classroom as a regular feature of instruction. Epistemic conceptual change may also be a precondition for promoting strong and lasting conceptual change (Sinatra & Chinn, 2012).

It is important to note that this study demonstrated a sustained, relatively long-term, effect on conceptual understanding of scientific consensus about human-induced climate change. Many conceptual change studies show that knowledge change can be ephemeral, with students shifting back toward their original conceptions over time (e.g., Guzzetti, Snyder, Glass, & Gamas, 1993). Students in this study tested six months after instruction sustained their conceptions. This is impressive given the relatively short-term (90 min across two instructional days) nature of the intervention. Perhaps the combined influence of critical evaluation of evidence and plausibility reappraisal was important for the maintenance of the effect, an issue that warrants further investigation.

Another possibility is that—given the immense importance of the topic—students were motivated to keep thinking about the topic once its plausibility was established, resulting in memory consolidation. A third possibility is that epistemic conceptual change may also be a contributing factor here, resulting in students continuing to think critically with a greater depth of processing in the six months before the delayed posttest. A fourth possibility is that for this topic, the competing alternative conceptions were not as ingrained as with some other topics. This may be true to some degree, but one of the core alternative conceptions—that humans do not affect the climate—could be very ingrained from personal experience. Here again, the combined effect of critical evaluation, which may have resulted in plausibility reappraisal (and then to perhaps continued critical evaluation), could have been an important factor in sustaining conceptual change. These various factors need to be investigated in future research.

4.1. Instructional implications

Our findings indicate that with carefully crafted instruction students can coordinate evidence with theories in a mode of critical evaluation. Recently, the U.S. National Academies of Science, Engineering, and Medicine published a report providing a framework for the next generation of science education standards (National Research Council, 2012). The report states that coordination of evidence and theory through critical evaluation supports the learning of epistemic practices of scientists and engineers (National Research Council, 2012). With the framework calling for changing students' conceptual understanding of such epistemic practices, instruction that supports critical evaluation and plausibility reappraisals may support these standards.

Engaging students explicitly in considering and reappraising their plausibility judgments may also increase students' understanding of the nature of science. A key component to understanding the nature of science is the idea that scientific knowledge is tentative (Lederman, 1999). But of equal importance to knowing that scientific knowledge is tentative, students should also "be able to step back from evidence or an explanation and consider whether another interpretation of a particular finding is *plausible* with respect to existing scientific evidence and other knowledge that they hold with confidence" (National Research Council, 2007, p. 39, emphasis ours). Even if an existing theory is highly accepted, alternative theories should always be considered (and not prematurely rejected as implausible), and evidence collected and considered to properly appraise and reappraise the plausibility of competing theories. Explicit and conscious reappraisal of plausibility judgments may be beneficial to deepen understanding about the nature of scientific knowledge and how scientific knowledge develops over time. Whereas individual scientists may not always actively engage in plausibility reappraisal of the theoretical frameworks critical to their research agenda, the larger scientific community will evaluate major scientific theories and eventually dispel those that are deemed less plausible than competing theories (Lakatos, 1970). Instruction using explicit plausibility reappraisal might then facilitate understanding of how scientific knowledge develops. In fact, the new framework for science education specifically calls for instruction where students "come to appreciate that alternative interpretations of scientific evidence can occur, that such interpretations must be carefully scrutinized, and that *the plausibility of the supporting evidence must be considered*" (National Research Council, 2012, p. 251, emphasis ours).

4.2. Limitations and future research

All studies have limitations and ours is no exception. The biggest limitation relates to external validity; we do not know if we would get the same effects on a different topic, or with a different age group. Replications of our study in different contexts are vital. There are several other topics that have "plausibility gaps," such as human evolution and other non-controversial but non-intuitive concepts such as the atomic theory of matter, where seemingly solid materials are made from unseen tiny particles and are mostly empty space. For topics with a large plausibility gap, instruction designed to promote plausibility reappraisal may be a key pedagogical strategy.

Another possible limitation to the study is the approximately 37% of the students who failed to return signed parental consent forms. Although very unlikely, it is possible that that lack of parental consent indicates potential bias of the results toward those who are more accepting of the climate change message. Based on the lead author's discussion with the participants' teachers and their past poor behavior in returning material from home, we believe that lack of parental consent was mainly attributable to students' carelessness. In addition, we should note that despite a potential bias, classes were randomly assigned to the treatment and comparison conditions. Gains shown by the treatment condition and lack of gains shown by the comparison condition demonstrate that instruction promoting critical evaluation (e.g., the model-evidence link diagram) may facilitate plausibility reappraisal and conceptual change and drastically eliminate any potential bias effects.

We acknowledge that inclusion of a true control group (i.e., a group receiving no climate change instruction) would have strengthened our conclusions. However, the realities of conducting research in authentic classroom settings make obtaining true control conditions very difficult. Another study limitation is that we did not collect data directly related to student–teacher interactions. Future research examining these interactions could provide important information about the broader effectiveness of using critical evaluation to promote plausibility reappraisal.

Future research should also extend our understanding of how other types of instruction promoting critical evaluation might lead to plausibility reappraisal and conceptual change. For example, Nussbaum and Edwards (2011) have shown that critical questions can be used to increase students' abilities to successfully evaluate arguments. Critical questions—such as "What is the likelihood?" and "How do scientists know?"—may enable students to evaluate connections between evidence and scientific models and potentially impact plausibility reappraisals, although more research in this area is warranted. Furthermore, incorporating collaborative argumentation into instruction may allow for greater elaboration and evaluation when explicitly considering both judgments based on plausibilistic reasoning as well as more precise probabilistic reasoning (Nussbaum, 2011). However, argumentation interventions may have a different impact if students reason in a biased and motivated fashion (Mercier & Sperber, 2011).

We also need to know how motivational factors may be related to the degree of evaluation that occurs in both the initial plausibility judgments and potential plausibility reappraisal. For example, Sinatra and Taasoobshirazi (2011) describe the process of intentional conception change, where "motivation drives the cognition and metacognition needed for conceptual change" (p. 209). With intentional conceptual change, individuals have the goal of examining incoming information in comparison to their background knowledge and evaluating the need for knowledge reconstruction. Research into whether explicit use of plausibility judgments may facilitate such a goal-directed comparison could provide greater understanding of the self-regulatory skills that promote conceptual change. In turn, this could help us better understand the interaction of learner and message characteristic as postulated by Dole and Sinatra's (1998) Cognitive Reconstruction of Knowledge Model (CRKM).

4.3. Conclusion

This study expands our understanding of conceptual change, and specifically helps support the idea that experiencing conceptual change about human-induced climate change is neither a simple matter of debunking non-scientific positions nor just learning about the several lines of evidence that support the scientific model. Rather, moving toward the scientifically accepted conception that humans are altering Earth's climate may well require connecting evidences to alternative models, evaluating the strength of these connections with respect to each alternative, and explicitly reappraising the plausibility of each alternative. Doing so in an instructional setting may seem counterintuitive to those that are involved in the climate change debate on a daily basis (given that virtually all climate scientists endorse human-induced climate change; Doran & Zimmerman, 2009), but individuals who are committed to developing a citizenry that is climate-literate must be open to the notion that discussing alternative explanations may lead to greater awareness and understanding of the science. Such literacy is critical to developing a society that characteristically exhibits scientific habits of mind and is equipped to deal with future challenges in a way that is beneficial to the global community.

Acknowledgments

This study was completed as part of the first author's doctoral dissertation while at the University of Nevada, Las Vegas, with the second and third authors serving as dissertation committee co-chairs. The authors would like to thank the other committee members, Dr. Gregory Schraw and Dr. Matthew Lachniet, both from the University of Nevada, Las Vegas. We would also like to thank the four teachers who opened their classrooms for this study, Ms. Gay Golden, Ms. Brenda Serber, Mr. Jose Trujillo, and Ms. Heather Torrey, as well as their students who participated in this study. Finally we greatly appreciate Dr. Janelle Bailey from University of Nevada, Las Vegas, and Dr. Michael Smith of Temple University, who provided valuable feedback on this research.

Appendix 1. Human-induced climate change (HICCK) instrument

Below are statements about climate change. Rate the degree to which you think that *climate scientists* agree with these statements.

(Note that items with an asterisk directly relate to alternative conceptions about the causes of climate change summarized by Choi et al., 2010. These asterisks were not included in the version used by participants.)

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. The Sun is the main source of energy for Earth's climate.	1	2	3	4	5
2. Humans has very little effect on Earth's climate.	1	2	3	4	5
3. We cannot know about ancient climate change.	1	2	3	4	5
4. Earth's climate has probably changed little in the past.	1	2	3	4	5

(*continued*)

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
5. The Sun's brightness is one way to measure solar activity.	1	2	3	4	5
6. Sunspot number is related to solar activity.	1	2	3	4	5
7. Greenhouse gases make up less than 1% of Earth's atmosphere.	1	2	3	4	5
8. Burning of fossil fuels produces greenhouse gases.	1	2	3	4	5
9. Humans produce billions of tons of greenhouse gases each year.	1	2	3	4	5
10. Humans are reducing the amount of fossil fuels they burn.	1	2	3	4	5
11. Greenhouse gas levels are increasing in the atmosphere.	1	2	3	4	5
12. Greenhouse gases absorb some of the energy emitted by Earth's surface.	1	2	3	4	5
13. Earth's climate is currently changing.	1	2	3	4	5
14. Humans are behind the cause of Earth's current climate change.	1	2	3	4	5
15. Earth's climate is not currently changing.	1	2	3	4	5
16. Current climate change is caused by human activities.	1	2	3	4	5
17. Current climate change is caused by an increase in the Sun's energy.*	1	2	3	4	5
18. Current climate change is caused by the ozone hole.*	1	2	3	4	5
19. Current climate change is caused by changes in Earth's orbit around the Sun.*	1	2	3	4	5
20. Current climate change is caused by volcanic eruptions.*	1	2	3	4	5
21. Current climate change is caused by increasing dust in the atmosphere.*	1	2	3	4	5
22. Future climate change may be slowed by reducing greenhouse gas emissions.	1	2	3	4	5
23. Humans cannot reduce future climate change.	1	2	3	4	5
24. Satellites do not provide evidence that humans are changing Earth's climate.	1	2	3	4	5
25. Earth's average temperature has increased over the past 100 years. This is evidence of climate change.	1	2	3	4	5
26. Average sea level is increasing. This is evidence of climate change.	1	2	3	4	5
27. Most of the world's glaciers are decreasing in size. This is evidence of climate change.	1	2	3	4	5

References

Boyes, E., & Stanisstreet, M. (1993). The greenhouse effect—children's perception of causes, consequences and cures. *International Journal of Science Education, 15*, 531–552. http://dx.doi.org/10.1080/0950069930150507.

Bråten, I., Strømsø, H. I., & Salmerón, M. A. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21*, 180–192. http://dx.doi.org/10.1016/j.learninstruc.2010.02.002.

Cavagnetto, A., & Hand, B. (2012). The importance of embedding argument within science classrooms. In M. S. Khine (Ed.), *Perspectives in scientific argumentation* (pp. 39–53). Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-94-007-2470-9_3.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: why some misconceptions are robust. *The Journal of Learning Sciences, 14*, 161–199. http://dx.doi.org/10.1207/s15327809jls1402_1.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: case studies in science classrooms. *Journal of the Learning Sciences, 19*, 230–284. http://dx.doi.org/10.1080/10508400903530036.

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science education. *Review of Educational Research, 63*, 1–49. http://dx.doi.org/10.3102/00346543063001001.

Chinn, C. A., & Buckland, L. A. (2012). Model-based instruction: fostering change in evolutionary conceptions and in epistemic practices. In K. S. Rosengren, E. M. Evans, S. Brem, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 211–232). New York: Oxford University Press.

Chinn, C. A., Duschl, R. A., Golan Duncan, R., Buckland, L. A., & Pluta, W. J. (2008). A microgenetic classroom study of learning to reason scientifically through modeling and argumentation. In. *Learning in the disciplines: Proceedings of the 8th international conference of the learning sciences (ICLS 2008): Vol. 3. Short papers, symposia, and selected abstracts*. Utrecht, the Netherlands: International Society of the Learning Sciences.

Choi, S., Niyogi, D., Shepardson, D. P., & Charusombat, U. (2010). Do earth and environmental science textbooks promote middle and high school students' conceptual development about climate change?: textbooks' consideration of students' misconceptions. *Bulletin of the American Meteorological Society, 91*, 889–898. http://dx.doi.org/10.1175/2009BAMS2625.1.

Clark-Carter, D. (2010). *Quantitative psychological research*. New York, NY: Psychology Press.

Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science, 30*, 95–120.

Cook, J. (2010). *Solar activity and climate: Is the sun causing global warming?* Skeptical Science. Retrieved from. http://www.skepticalscience.com.

Cress, U. (2008). The need for considering multilevel analysis in CSCL research: an appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning, 3*, 69–84. http://dx.doi.org/10.1007/s11412-007-9032-2.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*, 109–128. http://dx.doi.org/10.1207/s15326985ep3302&3_5.

Doran, P. T., & Zimmerman, M. K. (2009). Examining the scientific consensus on climate change. *EOS Transactions, 90*, 22–23. http://dx.doi.org/10.1029/2009EO030002.

Duschl, R. A. (2008). Science education in three-part harmony: balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*, 268–291. http://dx.doi.org/10.3102/0091732X07309371.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach, or Spearman–Brown? *International Journal of Public Health* http://dx.doi.org/10.1007/s00038-012-0416-3. Advance online publication.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*, 5–24. http://dx.doi.org/10.1037//0022-3514.71.1.5.

Ellis, B. (2009). NASA study shows sun responsible for planet warming. *Dakota Voice*. http://www.dakotavoice.com/2009/06/nasa-study-shows-sun-responsible-for-planet-warming/. Retrieved on 04.02.11.

George, D., & Mallery, P. (2009). *SPSS for Windows step by step: A simple guide and reference: 16.0 Update*. Boston, MA: Pearson Education, Inc.

Gowda, M. V. R., Fox, J. C., & Magelky, R. D. (1997). Students' understanding of climate change: insights for scientists and educators. *Bulletin of the American Meteorological Society, 78*, 2232–2240.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: a comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*, 116–159. http://dx.doi.org/10.2307/747886.

Halpern, D. F. (2007). Is intelligence critical thinking? Why we need a new definition for intelligence. In P. Kyllonen, I. Stankov, & R. D. Roberts (Eds.), *Extending intelligence: Enhancement and new constructs* (pp. 349–370). Mahwah, NJ: Erlbaum Associates, Inc.

Intergovernmental Panel on Climate Change. (2007). *Climate change 2007: Synthesis report—Summary for policymakers*. Geneva, Switzerland: World Meteorological Organization.

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology, 88*, 260–271. http://dx.doi.org/10.1037//0022-0663.88.2.260.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). Cambridge, England: Cambridge University Press.

Lakatos, I. (1970). History of science and its rational reconstructions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1970*, 91–136. http://dx.doi.org/10.1007/978-94-010-3142-4_7.

Lederman, N. G. (1999). Teachers' understanding of the nature of science and classroom practice: factors that facilitate or impede the relationship. *Journal of Research in Science Teaching, 36*, ,916–929. http://dx.doi.org/10.1002/(SICI)1098-2736(199910)36:8<916::AID-TEA2>3.0.CO;2-A.

Leiserowitz, A., & Smith, N. (2010). *Knowledge of climate change across global warming's six Americas*. Yale University. New Haven, CT: Yale Project on Climate Change Communication.

Lockwood, M. (2010). Solar change and climate: an update in the light of the current exceptional solar minimum. *Proceedings of the Royal Society of Astronomy*. http://dx.doi.org/10.1098/rspa.2009.0519. Advance online publication.

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. Plausibility judgments in conceptual change and epistemic cognition, in review.

Lombardi, D., & Sinatra, G. M. (2012). College students' perceptions about the plausibility of human-induced climate change. *Research in Science Education, 42*, 201–217. http://dx.doi.org/10.1007/s11165-010-9196-z.

Lombardi, D., & Sinatra, G. M. (2013). Emotions when teaching about human-induced climate change. *International Journal of Science Education, 35*, 167–191. http://dx.doi.org/10.1080/09500693.2012.738372.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*, 153–191. http://dx.doi.org/10.1207/s15327809jls1502_1.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111. http://dx.doi.org/10.1017/S0140525X10000968.

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

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: preface and literature review. *Contemporary Educational Psychology, 33*, 345–359. http://dx.doi.org/10.1016/j.cedpsych.2008.06.001.

Nussbaum, E. M. (2011). Argumentation, dialogue theory, and probability modeling: alternative frameworks for argumentation research in education. *Educational Psychologist, 46*, 84–106. http://dx.doi.org/10.1080/00461520.2011.558816.

Nussbaum, E. M., & Edwards, O. V. (2011). Critical questions and argument stratagem: a framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20*, 443–488. http://dx.doi.org/10.1080/10508406.2011.564567.

Österlind, K. (2005). Concept formation in environmental education: 14-year olds' work on the intensified greenhouse effect and the depletion of the ozone layer. *International Journal of Science Education, 27*, 891–908. http://dx.doi.org/10.1080/09500690500038264.

Papadimitriou, V. (2004). Prospective primary teachers' understanding of climate change, greenhouse effect, and ozone layer depletion. *Journal of Science Education and Technology, 13*, 299–307. http://dx.doi.org/10.1023/B: JOST.0000031268.72848.6d.

Penuel, W. R., & Gallagher, L. P. (2009). Preparing teachers to design instruction for deep understanding in middle school earth science. *Journal of the Learning Sciences, 18*, 461–508. http://dx.doi.org/10.1080/10508400903191904.

Pintrich, P. R., Marx, R. W., & Boyle, R. B. (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/10.2307/1170472.

Popper, K. (1963). *Conjectures and refutations*. London, England: Routledge. http://dx.doi.org/10.1063/1.3050617.

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.

Priest, E., Lockwood, M., Solanki, S., & Wolfendale, A. (2007). Does the sun affect earth's climate? *Astronomy and Geophysics, 48*, 3–7 http://dx.doi.org/10.1111/j.1468-4004.2007.48307.x.

Pruneau, D., Gravel, H., Courque, W., & Langis, J. (2003). Experimentation with a socio-constructivist process for climate change education. *Environmental Education Research, 9*, 429–446. http://dx.doi.org/10.1080/1350462032000126096.

Rescher, N. (1976). *Plausible reasoning: An introduction to the theory and practice of plausibilistic inference*. Amsterdam, The Netherlands: Van Gorcum Ltd.

Shepardson, D. P., Choi, S., Niyogi, D., & Charusombat, U. (2011). Seventh grade students' mental models of the greenhouse effect. *Environmental Education Research, 17*, 1–17. http://dx.doi.org/10.1080/13504620903564549.

Sinatra, G. M., & Chinn, C. (2012). Thinking and reasoning in science: promoting epistemic conceptual change. In K. Harris, S. Graham (Series Eds.) & K. Harris, C. B. McCormick, G. M. Sinatra, & J. Sweller (Vol. Eds.), *APA educational psychology handbook series: Vol. 1. Critical theories and models of learning and development relevant to learning and teaching* (pp. 257–282). APA Publications. http://dx.doi.org/10.1037/13275-011.

Sinatra, G. M., & Taasoobshirazi, G. (2011). Intentional conceptual change: the self-regulation of science learning. In B. Zimmerman, & D. Shunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 203–216). New York, NY: Routledge.

Smith, M. J., Southard, J. B., & Mably, C. (2002). *Investigating earth systems: Climate and weather: Teacher's edition*. Armonk, New York: It's About Time, Inc.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Solanki, S. K., Usokin, I. G., Kromer, B., Schüssler, M., & Beer, J. (2004). Unusual activity of the sun during recent decades compared to the previous 11,000 years. *Nature, 431*, 1084–1087. http://dx.doi.org/10.1038/nature02995.

Southerland, S. A., Sinatra, G. M., & Matthews, M. R. (2001). Belief, knowledge, and science education. *Educational Psychology Review, 13*, 325–351.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: the relationship between memory for sources and text comprehension. *Learning and Instruction, 20*, 192–204. http://dx.doi.org/10.1016/j.learninstruc.2009.02.001.

Szu, E., & Osborne, J. (2012). Scientific reasoning and argumentation from a Bayesian perspective. In M. S. Khine (Ed.), *Perspectives in scientific argumentation* (pp. 55–71). Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-94-007-2470-9_4.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.

Treagust, D. F., & Duit, R. (2008). Conceptual change: a discussion of theoretical, methodological, and practical challenges for science education. *Cultural Studies of Science Education, 3*, 297–328. http://dx.doi.org/10.1007/s11422-008-9090-4.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69.

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.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality & Social Psychology, 67*(6), 1049–1062. http://dx.doi.org/10.1037//0022-3514.67.6.1049.