Students’ evaluations about climate change

Doug Lombardi, Carol B. Brandt, Elliot S. Bickel and Colin Burg

Department of Teaching & Learning, Temple University, Philadelphia, PA, USA

ABSTRACT
Scientists regularly evaluate alternative explanations of phenomena and solutions to problems. Students should similarly engage in critical evaluation when learning about scientific and engineering topics. However, students do not often demonstrate sophisticated evaluation skills in the classroom. The purpose of the present study was to investigate middle school students’ evaluations when confronted with alternative explanations of the complex and controversial topic of climate change. Through a qualitative analysis, we determined that students demonstrated four distinct categories of evaluation when writing about the connections between evidence and alternative explanations of climate change: (a) erroneous evaluation, (b) descriptive evaluation, (c) relational evaluation, and (d) critical evaluation. These categories represent different types of evaluation quality. A quantitative analysis revealed that types of evaluation, along with plausibility perceptions about the alternative explanations, were significant predictors of postinstructional knowledge about scientific principles underlying the climate change phenomenon. Specifically, more robust evaluations and greater plausibility toward the scientifically accepted model of human-induced climate change predicted greater knowledge. These findings demonstrate that instruction promoting critical evaluation and plausibility appraisal may promote greater understanding of socio-scientific topics and increased use of scientific thinking when considering alternative explanations, as is called for by recent science education reform efforts.

Introduction and purpose

Recent educational reform efforts stress the importance of students engaging in evaluation in the science classroom. For example, in the US, A Framework for K-12 Science Education places evaluation at the nexus between the scientific activities of ‘investigating’ and ‘developing explanations and solutions’ (National Research Council, 2012, p. 45). This framework further states that the practice of evaluation requires ‘critical thinking’ in both ‘developing and refining an idea (an explanation or a design) or in conducting an investigation’ (p. 46). This process of critical evaluation would require individuals to distinguish between and coordinate evidence with scientific explanations (e.g. hypotheses and theories ‘of how phenomena unfold the way they do’; Braaten & Windschitl, 2011, p. 641).
curriculum reform in South Africa has similarly stressed the importance of students evaluating the quality of arguments through coordination of evidence and explanation (Erduran & Msimanga, 2014). Such practice demonstrates mature scientific and reflective thinking (King & Kitchener, 2004; Kuhn & Pearsall, 2000). However, many students exiting high school do not express advanced reasoning associated with critical evaluation (Erduran & Msimanga, 2014; King & Kitchener, 2004).

Engaging in critical evaluation is particularly important for complex and/or controversial socio-scientific phenomena, such as climate change, where a potential gap exists between what laypersons (e.g. students) and scientists find plausible (Lombardi, Sinatra, & Nussbaum, 2013). Developing a deep understanding about climate change may be difficult not only because the underlying scientific principles are complex, but also because understanding why scientists think that Earth’s global climate is changing is also complex. Students also need to engage in epistemic cognitive processes that reflect scientific reasoning used to connect evidence and explanations to gain better understanding (Lombardi, Nussbaum, & Sinatra, 2016; Sinatra & Chinn, 2011). This ability to conduct critical evaluation is also integral to collaborative argumentation in an intellectual community, especially where rival interpretations vie for acceptance.

The purpose of the present study was to investigate middle school students’ evaluations of competing explanations about the cause of current climate change. One of these explanations was the scientifically accepted idea that human activities are the primary cause of recent increases in global average temperatures, sea-level rise, and ice-sheet melting (Doran & Zimmerman, 2009). The other explanation, popularized by climate change skeptics, was that current climate change is caused by an increasing amount of solar energy received by Earth (Cook, 2010; Ellis, 2009). Middle school students engaged with these two alternative explanations using a model-evidence link (MEL) diagram during classroom instruction. Earlier, Lombardi et al. (2013) reported that instruction using the climate change MEL resulted in significant shifts in both students’ plausibility and knowledge of climate change toward the scientifically accepted explanation. Knowledge gains were maintained up to six months after instruction. In the present study, we examined the underlying evaluative mechanisms that led to these plausibility and knowledge shifts. Specifically, our research questions were:

(1) What types of evaluations do middle school students use when considering alternative explanations of climate change during classroom instruction?
(2) How do evaluations about these alternative explanations relate to plausibility judgments and knowledge about climate change after instruction?

Using a mixed methods approach (i.e. by qualitative and quantitative analyses), we investigated middle school students’ evaluations when confronted with alternative explanations of the complex and controversial topic of climate change. After grounding our research in the literature on the role of critical evaluation in scientific reasoning and plausibility perceptions, we describe a qualitative analysis that documents students’ demonstrated distinct categories of evaluation when writing explanations about lines of evidence and alternative explanations of climate change. We follow by presenting quantitative results that revealed types of evaluation. These findings, along with plausibility perceptions about alternative explanations, were significant predictors of postinstructional knowledge about scientific
principles underlying climate change. These findings point to ways that critical evaluation – as part of science instruction – can facilitate plausibility appraisal, greater understanding of socio-scientific topics, and increased scientific thinking. As a result of this analysis, we contend that engaging in critical evaluation to reappraise plausibility is a skill that can assist students as they develop the ability to reason scientifically, as well as understand how scientific knowledge is constructed in collaborative discussion.

**Theoretical framework**

Our perspective on critical evaluation in scientific reasoning and plausibility perceptions draws upon developmental psychology, educational psychology, and science education research. Central to our conceptual framework is the notion that an evaluative level of epistemological understanding—a fundamental component of critical thinking – involves judgments about the quality of explanations based on ‘criteria of argument and evidence’ (Kuhn, 1999, p. 23). For example, ‘good models are consistent with empirical evidence’ (Pluta, Chinn, & Duncan, 2011, p. 486) is an epistemic criterion used by scientists to evaluate an explanation’s validity (Duschl, Schweingruber, & Shouse, 2007). The MEL diagram used in this study is an instructional scaffold built upon the idea of epistemic criteria, and is designed to make the scientific practice of critical evaluation explicit through model-based reasoning and argumentation (Chinn & Buckland, 2012). The MEL specifically helps students to weigh the connections between evidence and alternative models, thereby facilitating evaluation and ‘understanding that growth in scientific knowledge is a dynamic process’ (Erduran & Dagher, 2014, p. 126). Classroom use of epistemic criteria to evaluate scientific knowledge is an important dimension of collaborative argumentation (Jiménez-Aleixandre & Erduran, 2007). Furthermore, this study tests a theoretical model that views critical evaluation as a central component in the dynamic process of plausibility reappraisal and knowledge construction (Lombardi et al., 2016). This process of plausibility reappraisal may be a particularly relevant factor when scientific topics are controversial and/or complex (e.g. climate change). The following subsections provide more details on the interconnections of critical evaluation, scientific reasoning, and plausibility to help support our reasons for conducting the present study.

**Critical evaluation**

Students may be naturally curious about scientific topics, but they are not necessarily evaluative as they consider hypotheses and theories. Critical evaluation in science learning situations can involve judgments about the relationship between evidence and alternative explanations of a particular phenomenon (McNeill, Lizotte, Krajcik, & Marx, 2006). Through critical evaluation, an individual seeks to weigh the strengths and weaknesses in the connection between evidence and explanations. Mere critique is not sufficient because critical evaluation involves gauging how well evidence potentially supports both an explanation (e.g. an argument, a scientific model) and its plausible alternatives (e.g. a counterargument, a contrary hypothesis). In this way, critical evaluation embraces the criterion of falsifiability, where evidence may invalidate an explanation in favor of another alternative (Popper, 1963; Stanovich, 2007). Whereas individual scientists
might not adhere to the falsifiability criterion, the scientific community ultimately eliminates hypotheses and theories with demonstrated evidentiary failures (Lakatos, 1970).

Critical evaluation demands that students be reflective about the process of knowledge construction (Mason, Ariasi, & Boldrin, 2011). Such reflection may be facilitated when students model practices used by scientific experts (Duschl et al., 2007). Students who engage in critical evaluation understand that scientific knowledge emerges from collaborative argumentation, which is a constructive and social process where individuals compare, critique, and revise ideas (Nussbaum, 2008). Chin and Osborne (2010) suggest that critical evaluation is stimulated by argumentative discourse activities, where students challenge each other’s thinking through questions about the strength of evidence and explanation connections. Collaborative argumentation is different from adversarial argumentation, where opponents attempt to reduce one another’s viewpoint to a point of uselessness. Individual scientists may engage in adversarial argumentation; after all, scientists are human too. But as a community, science thrives due to collaborative argumentation, which is an inherently constructive process (Osborne, 2010).

Argument construction, however, does not necessarily promote greater critical evaluation. A study conducted by Nussbaum and Kardash (2005) showed that when students were given a persuasion goal, they were less critical. Trying to persuade led to generation of fewer counterarguments and one-sided thinking. Nussbaum and Kardash also found a connection with the intensity of students’ beliefs about a topic and their ability to generate counterarguments, where more extreme beliefs led to fewer counterarguments. Because students may not naturally be critically reflective when engaging in collaborative argument, they may need instructional scaffolds to evaluate the quality of explanations (Nussbaum & Edwards, 2011). A promising scaffold that may help students develop deeper levels of evaluative thinking is the MEL diagram, which assists students in effectively coordinating evidence with scientific explanations (Chinn & Buckland, 2012). The MEL facilitates evaluation and helps students differentiate between evidence and scientific explanations – a scientific reasoning skill in which students often have difficulty (Duschl & Grandy, 2011; Kuhn & Pearsall, 2000).

**Scientific reasoning and critical evaluation**

Scientific reasoning involves critical evaluations about the strength of connections between lines of evidence and alternative explanations of phenomenon. Our characterization of reasoning follows Johnson-Laird’s (1983) idea: individuals construct and use mental models to reason about and develop explanations. From such a perspective, scientific reasoning occurs when individuals link observational evidence of reality to explanatory models of how the universe functions (e.g. hypotheses and theories; Erduran & Dagher, 2014). Nersessian (1999) argues that such model-based reasoning, whether experimentally or theoretical based, is how scientific concepts are formed and changed over time. These changes are facilitated in the scientific community through argumentation, in which knowledge is evaluated through epistemic criteria (Jiménez-Aleixandre & Erduran, 2007). Furthermore, model-based reasoning involves the use of ‘analogies, imagistic representations, and thought experiments work[ing] hand in hand, first with experimental investigations and then with mathematical analyses’ (Nersessian, 2008, p. x). Therefore, critical evaluation is employed in model-based reasoning when alternative explanations
(formed through analogies and representations) are gauged via evidence gathered through experiment and analysis.

Yet students do not necessarily engage fully in model-based reasoning in the science classroom. Driver, Leach, Millar, and Scott (1996) formulated three qualitatively different reasoning types exhibited by students when engaged in science instruction. The first type is phenomenon-based reasoning, where students make little or no distinction between scientific evidence and explanation. The second type is relation-based reasoning, in which students connect evidence and explanation through simple correlational reasoning (e.g. correlation implies causation). The third type is model-based reasoning, where students weigh the strength of the evidence supporting an explanation, and in some circumstances, weigh the strength of evidence supporting multiple, alternative explanations. This final circumstance (i.e. weighing of alternatives) would be a reasoning situation in which an individual would be critically evaluative. Students have difficulty engaging in model-based reasoning. However, instructional scaffolds that promote critical evaluation of alternatives during argumentation by employing epistemic criteria (e.g. the MEL) may facilitate engagement and deeper understanding (Erduran & Dagher, 2014; Jiménez-Aleixandre & Erduran, 2007). This may especially be the case for topics – such as global climate change – where scientists greatly rely on models and modeling of complex systems.

**Plausibility and critical evaluation**

Students encountering science topics in a school setting often possess existing mental representations that conflict with scientific understanding, and often these naïve understandings seem more plausible than the correct conception (Dole & Sinatra, 1998; Posner, Strike, Hewson, & Gertzog, 1982). According to the Cognitive Reconstruction of Knowledge Model (CRKM; Dole & Sinatra, 1998), plausibility is one of four judgments that students make about an incoming message (the others being how comprehensible, compelling, and coherent the message is). These four message characteristics interact with learner characteristics (e.g. background knowledge, motivational factors, personal relevance, and cognitive dispositions) to influence the degree of engagement that a student has with a topic. Consequently, a higher level of engagement could promote deeper understanding. A recently developed theoretical model of plausibility judgments expands upon Dole and Sinatra’s CRKM by mapping out how plausibility may specifically interact with motivation, emotion, and epistemic dispositions – important factors that influence the degree of cognitive evaluation (Lombardi et al., 2016). This perspective describes how plausibility judgments may be most often formed through automatic cognitive processes. However, explicit instruction may also promote reappraisal of these implicit plausibility judgments. Such instruction may be particularly relevant for complex and abstract scientific topics (e.g. climate change), where a gap exists between what students and scientists find plausible.

Recent research has revealed that a plausibility gap exists for the topic of global climate change among middle school students (Lombardi et al., 2013), undergraduate students (Lombardi & Sinatra, 2012), and elementary and secondary science teachers (Lombardi & Sinatra, 2013). To address this gap, Lombardi et al. (2013) developed a MEL diagram for climate change and used this instructional scaffold in grade 7 Earth science classrooms. In this study, participants in the treatment group (using the climate change MEL diagram) experienced significant shifts in both plausibility and knowledge toward the scientifically
accepted model of human-induced climate change. Furthermore, these students retained knowledge gains six months after instruction. In comparison, grade 7 participants at the same school and taught by the same teachers did not experience plausibility or knowledge shifts when experiencing a curricular activity (Smith, Southard, & Mably, 2002) designed to promote scientific inquiry and deeper understanding of climate change. Although the comparison activity asked students to link evidence to explanations, the primary difference from the treatment activity (i.e. the MEL diagram) was that students did not weigh evidence between two competing models of climate change.

Lombardi et al. (2013) speculated that the students’ plausibility reappraisal – a skill that is important for understanding the development of scientific knowledge (Duschl et al., 2007; Hogan & Maglienti, 2001) – was related to the MEL’s ability to facilitate students’ critical evaluation. Plausibility reappraisal, in turn, may have promoted the students’ enduring knowledge gains (Erduran & Dagher, 2014).

Plausibility and argumentation

Plausibility judgments are central to Walton’s (2007) argumentation framework, which posits a dynamic relationship between various argumentation schemes, critical questions, answers or refutations, and abductive inferences (i.e. inference to the best explanation; Harman, 1965). Specifically, individuals gauge the relative plausibility of alternative explanations (e.g. an argument-counterargument) through the process of abduction (Walton, 2004). In educational research, argumentation interventions based on Walton’s framework have been tested, revealing promising results in promoting students’ critical evaluation (Nussbaum, 2011; Nussbaum & Edwards, 2011). However, in science education, Toulmin’s (1958) Argumentation Pattern is the default framework of choice (see, e.g. Christodoulou & Osborne, 2014; Gray & Kang, 2014; Kulatunga, Moog, & Lewis, 2013). Toulmin’s framework posits that arguments are constructed through six components: (a) claims, (b) grounds, (c) warrants, (d) backing for warrants, (e) rebuttals, and (f) modal qualifiers. As such, Toulmin’s framework provides little support for the reasoning processes involved in argument, specifically the critical thinking skills that are necessary to engage in collaborative and constructive discourse (Zohar, 2007). Although we acknowledge that Toulmin’s Argument Pattern in the study of argumentation discourse provides a model for students to justify their claims (Erduran, Simon, & Osborne, 2004), a plausibility-based approach may open up an understanding of students’ reasoning and thinking, as well as potential misunderstandings of key concepts. Therefore, the purpose of the present study is to examine this speculative claim in more detail, by specifically investigating the different types of evaluations demonstrated by students in an explanatory task. To examine students’ evaluations, we conducted a thorough analysis of their written explanations about the links students drew on their climate change MEL diagrams, a previously untapped data source.

Methods

Participants and setting

Middle school students from a large urban district in the Southwestern United States participated in the study. The school district involved in this study teaches about climate change.
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. For the present study, we only included students who provided both parental consent and self-assent, fully completed all study activities, and were part of the treatment group that used the MEL materials \((n=85)\). Of the 85 students who participated in the present study, 55 (65%) were Hispanic, 14 (16%) were White, 13 (15%) were African American, and 3 (4%) were Asian/Pacific Islander. Forty-five participants (53%) were male. Eight (9%) of the participants had individualized education plans, 18 (21%) had limited proficiency in the English language, and 40 (47%) were eligible for free or reduced-cost lunch. Again, the present study concerned a detailed analysis on the MEL explanatory task and did not include details on the other participants used in a previous comparative study that looked at overall effectiveness of the MEL intervention. However, note that the comparison group was of similar size \((n=81)\) and demographic composition to the participants in the present study (see Lombardi et al., 2013, for more details on the comparison group and associated results).

**Materials**

We conducted the study towards 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 minutes of instructional time total). Seven total classes were involved in the study (three different teachers were instructors for two classes each and one teacher was the instructor for one class). Details about the procedure and intervention follow the subsequent discussion of materials.

**Instructional scaffold**

Participants used the MEL diagram activity as the instructional intervention for the present study (see Figure 1 for a student-completed MEL). On the first page of the MEL, participants drew different types of arrows linking lines of evidence to the two alternative models of climate change (Model A: human-induced climate change and Model B: solar irradiance causing climate change). 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.

On the second page of the MEL activity, which was the focus of the present study, students completed the explanatory task. This task asked participants to select three (out of a possible eight) evidence-to-model links that they had made on their MEL diagram (i.e. the first page of the activity). In their explanations, participants identified each end of the link, with a line of evidence (numbered 1, 2, 3, or 4; see Figure 1) at one end and the model (A: human-induced or B: solar irradiance) at the other. Participants then circled their judgment about the weighting of a link’s strength between an evidence and a model (i.e. the evidence strongly supports the model, the evidence supports the model, the evidence has
nothing to do with the model, or the evidence contradicts the model). The participants also provided a justification for their weighting of a link’s strength, starting with the provided prompt ‘because.’ For example, a full explanation from one participant said, ‘Evidence #1 strongly supports Model A because atmospheric greenhouse gases have been rising for the past 50 years because of humans.’

Perceptions of model plausibility
Two items measured participants’ plausibility judgment of Model A (human-induced) and Model B (solar irradiance). 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. Participants completed these model ratings immediately prior to and
just after engaging in the MEL activity. For the purposes of the present study, we used only postinstructional ratings to examine the relationship between participants’ evaluations that occurred during the activity and their plausibility perceptions. 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 .70, which is at the threshold commonly considered acceptable (George & Mallery, 2009).

Knowledge of climate change
We used a 27-item instrument to measure participants’ human-induced climate change knowledge (HICCK; Lombardi et al., 2013) just prior to and immediately after instruction. This instrument measures conceptions about the current scientific consensus on climate change based on three sources: a recent study that surveyed American citizens on their understanding of scientific phenomena related to global warming (Leiserowitz & Smith, 2010); a 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. An example item states, ‘current climate change is caused by an increase in the Sun’s energy.’ 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. Reliability of the HICCK was at the acceptable threshold (Cronbach’s $\alpha$ = .69; George & Mallery, 2009). Similar to model plausibility ratings, we used only postinstructional knowledge scores in the present study to examine the relationship between the activity and knowledge.

Procedures
Prior to the instructional activity, participants completed the HICCK instrument and climate change model ratings of plausibility. Before completing the preinstructional ratings of model plausibility, participants read a short introduction to the two models and a statement defining plausibility. This short introduction exposed participants to the notion of plausibility. Each teacher conducted a short discussion to help the students clarify any misunderstandings about the models and plausibility. Each teacher and class group participated in this short discussion for about the same amount of time.

Participants then engaged in a pre-activity 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 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 Figure 1).

After making their initial rankings, participants read a short paragraph discussing falsifiability, and specifically, how evidence that contradicts an idea has a large influence on how scientific knowledge changes (see Appendix). 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 does have the greatest weight in changing scientists’ plausibility judgments. Although measuring the degree of change between the initial and final rankings was beyond the scope of the present study, during pilot testing the first author observed that many, but not all, would change their most important connection from strongly supports to contradicts.

During the MEL activity, participants individually read short expository texts discussing each piece of evidence (i.e. one page of text for each line of evidence). These pages also included graphs and figures. Teachers asked the students if they had any questions about the evidence texts, figures, and graphs to clear up any confusion or misunderstandings. 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 the written explanatory task (i.e. the part of the activity that is the focus of the present study), which allowed students to reflect on the arrows they drew on the MEL. Then participants rated each model’s plausibility individually (i.e. the same as they did during preinstruction). At the end of the learning activities, participants completed the HICCK for a second time.

**Qualitative data analysis**

We conducted a content analysis, which is a technique for systematically coding large amounts of text to create a small number of content categories (Stemler, 2001), to examine participants’ explanations. The lead, third, and fourth author independently read through the explanations multiple times. Between reading episodes, we compared coding results. After four iterations, we eventually focused our coding on the types of scientific reasoning exhibited by students during science instruction. These categories somewhat resembled those coined by Driver et al.’s (1996) types of scientific reasoning exhibited by students during science instruction. Driver et al.’s framework was based on students’ discourse during instruction and categorically divided all of the students’ scientific reasoning into three categories: phenomenon-based, relation-based, and model-based. Like Driver et al. (1996), we used these designations as our analytical categories. Additionally, we attended to indicators revealing participants’ degree of elaboration (i.e. ‘issue-relevant arguments’ contained in an explanation; Petty & Cacioppo, 1986, p. 128). Elaboration exists on various levels, with high elaboration associated with deep cognitive engagement and low elaboration with superficial cognitive engagement (Dole & Sinatra, 1998). Thus the final content analysis revealed that explanations fell into four well-defined categories of evaluation, which reflect both participants’ scientific reasoning and elaboration in their explanations of evidence-to-model links. These four categories, discussed in more detail below, represent a synthesis of and metamorphosis from Driver et al.’s (1996) and Dole and Sinatra’s (1998) frameworks.
**Category 1: Erroneous evaluation**

Many participants wrote incorrect explanations about evidence-to-model links, a category not addressed in Driver et al.’s (1996) types of reasoning. For example, one participant said that Evidence #2 strongly supports Model B because the evidence ‘talks about the Sun’s energy and the temperature rising and Model B talks about energy released from the Sun.’ The participant was clearly incorrect because Evidence #2, which reveals that solar activity has been decreasing, contradicts Model B, which states that our current climate change is caused by increasing amounts of energy released from the Sun. Although any response indicates some level of thought by the participant, erroneous model to evidence links imply that participants were not able to reach initial understandings necessary for substantive evaluation. Such erroneous evaluations could have resulted from lack of attention to the evidence and/or model text. Alternatively, erroneous evaluations may have emerged from a psychological response where ‘students sometimes ignore information in science texts that contradicts their existing schemas’ (Chinn & Brewer, 1993. p. 5).

Given that a clear category of erroneous evaluation emerged from the content analysis, we needed to look carefully at whether a given link between a line of evidence and an explanatory model is correct or incorrect (Table 1). Note that part of correctness is based on participants’ judgments about weight of a link’s strength (i.e. strongly supports, supports, has nothing to do with, or contradicts). The table combines the weights of ‘strongly supports’ and ‘supports,’ because from the perspective of correctness, it was not possible to differentiate between these two. We determined correct and incorrect responses based solely on the information provided in the evidence and the cause/effect statement made in a model. Although someone with a sufficient amount of background knowledge (i.e. an expert in climate science) could argue that other correct options exist, such nuances are beyond the level of these middle school participants, who were clearly novices in the area of climate science. Table 1 also shows if correct links are weak or strong, which reflects other types of elaboration and reasoning, which we discuss in the next three categories (see below).

**Category 2: Descriptive evaluation**

Often, explanations reflected correct explanations that were associated with a weak weighting of the connection between a line of evidence and an explanatory model. Consider, for example, a response that stated, ‘Evidence #3 has nothing do with Model B

<table>
<thead>
<tr>
<th>Evidence-to-model link</th>
<th>Link weight</th>
<th>Strongly supports/supports</th>
<th>Has nothing to do with</th>
<th>Contradicts</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1_MA</td>
<td>C+</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E1_MB</td>
<td>I</td>
<td>C–</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E2_MA</td>
<td>I</td>
<td>C–</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E2_MB</td>
<td>I</td>
<td>I</td>
<td>C+</td>
<td></td>
</tr>
<tr>
<td>E3_MA</td>
<td>C+</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E3_MB</td>
<td>I</td>
<td>C–</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E4_MA</td>
<td>I</td>
<td>C–</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>E4_MB</td>
<td>C+</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

Note: In the table, evidence-to-model links are coded based on the evidence number (1, 2, 3, or 4) and model (A or B) at each end of the link (e.g. E1_MA therefore shows the link from evidence #1 to Model A).
because the evidence is about satellites and greenhouses, and Model B is about energy released from the Sun.’ Many of the participants’ explanations discussed how certain evidence has nothing to do with a particular model. These explanations were often correct, but only tacitly so. In other words, indicating the evidence had nothing to do with a particular model is often a weak level of cognitive processing, with little or no elaboration (i.e. little or no issue-relevant arguments made in the explanation and ‘superficial or heuristic processing of information’; Dole & Sinatra, 1998. p. 121). These types of explanations share some similarity to Chinn and Brewer’s (1993) psychological response of excluding the data from the domain of the theory. When data are excluded, ‘they obviously do not lead to any theory change’ (Chinn & Brewer, 1993, p. 8) nor deep understanding about the topic.

Descriptive evaluations often demonstrated phenomenon-based reasoning as well, in that students made no distinction between a particular line of evidence and an explanatory model. For example, one participant wrote that Evidence #1 strongly supports Model A because ‘they are both talking about gasses [sic] and greenhouse gases.’ This student’s reasoning is based on the similarity between the text in Evidence #1 and Model A, with no clear distinction made between the two. Similarly, another participant wrote, ‘they both [i.e. the evidence and the model] talk about the sun affecting climate.’ These explanations make no attempt to address the scientific processes discussed in the evidences or to relate these pieces of evidence to the plausibility of the model.

**Category 3: Relational evaluation**

Many of the participants correctly discussed links that had strong connections (i.e. contradicts, supports, or strongly supports) to a particular model. These were signs of a deeper level of processing by indicating commitment (i.e. taking a definite positional stance), which in some cases could lead to a greater cognitive engagement (Dole & Sinatra, 1998). However, relational evaluation refers to many cases where, despite taking a commitment, written explanations were relatively superficial (i.e. lacking depth of analysis). Such a response was given by one participant who said, ‘Evidence #1 strongly supports Model A because Evidence #1 talks about greenhouse gases just like Model A.’ Despite noting scientific continuity between the evidence and model, this participant did not thoroughly explain the nature of the relationship between evidence and model. At best, correct and strong links with superficial explanations reflect a low to moderate level of elaboration because, even though participants are making meaningful connections between evidence and a model, they are still not thinking beyond surface details. These superficial connections may be akin to peripheral cues that are associated with low cognitive engagement. (Dole & Sinatra, 1998).

These types of participant explanations also showed relation-based reasoning (i.e. the second category in Driver et al.’s, 1996 framework). In the previous example, there was no attempt to explain how the evidence actually affects plausibility of the model. Instead, the participant notes similarity between the relevant components and considers this sufficient to give the link a value of ‘strongly supports.’ When students engage in relation-based reasoning, they are associating evidence to explanatory models by making a clearer distinction between lines of evidence and explanatory models, but are still focusing on similarity in text. As Driver et al. (1996) note, this type of relational reasoning shows that some students think that direct correlation implies causation (or lack of direct correlation implies lack of causation).
**Category 4: Critical evaluation**

Some participant explanations of strong evidence-to-model links expressed a greater degree of elaboration than was found in the others. For example, one participant indicated that, ‘Evidence #3 strongly supports Model A because the satellites are measuring energy being absorbed by greenhouse gases, which makes the Earth’s climate change.’ This participant provided an explanation about increasing energy absorbed by greenhouse gases as a mechanism for climate change, which in turn corresponded to Model A. This explanation demonstrates deeper thinking about the connection between lines of evidence and explanatory models. In these types of explanations, participants discussed distinctions between lines of evidence and explanatory models, as well as demonstrated more sophisticated types of coherence (potentially involving a nonlinear and/or discontinuous connection). Such explanations reflect model-based reasoning (Driver et al., 1996).

One participant wrote that Evidence #1 strongly supports Model A because ‘human activities have led to a greater release of greenhouse gases … Model A says that climate change is caused by increasing amounts of human activity.’ This student made a more sophisticated cause-and-effect relationship between human activities and climate change, with increased greenhouse gas emissions as the mediating variable. Likewise, another student wrote that Evidence #2 contradicts Model B because ‘Evidence #2 says that earth’s temperature continues to rise without the sun’s energy, but Model B says that earth’s temperature rises because of the sun’s energy.’ The student was able to clearly differentiate between evidence and explanatory model, and identify how the evidence contradicts the explanatory model. Explanations demonstrating critical evaluation could also concurrently examine the alternative models. For example one student wrote,

> As known, the conflict between the two thoughts [i.e., the models] would show the opposite, usually people would think that evidence 1 would conflict with B, but … my thoughts are from the increases about of sunspots, the sun gives out more energy, which makes it hot … starting gases from human activity, which also makes it hotter. Showing the two parts are together to make a large answer.

In this way, the student is weighing Evidence #1 to Model B (sun-induced climate change), but also concurrently considering Model A (human-induced climate change) in constructing the explanation.

**Types of evaluation rubric**

Table 2 shows a rubric for the four types of evaluation we identified in our qualitative content analysis. These four categories also represent a natural ordering of evaluation – from a low level of evaluation (erroneous, score = 1) to a low-moderate level of evaluation (descriptive, score = 2) to a moderate level of evaluation (relational, score = 3) to high level of evaluation (critical, score = 4).

**Quantitative results**

We specifically used rubric scores of types of evaluation in participants’ explanations, as well as scores on postinstructional measures of model plausibility and knowledge, to
examine the potential relationship between (a) student evaluations about alternative explanations of the causes of current global climate, and (b) plausibility judgments and (c) knowledge about climate change. Prior to presenting the results of statistical tests used to gauge this relationship, we first discuss the quantitative results of the scoring rubric used to measure participants’ type of evaluation in their explanations.

**Scoring rubric results for written explanations**

We scored participants’ explanations using the types of evaluation rubric (Table 2). As a reminder, the second page of the instructional activity asked each participant to provide three explanations (see Figure 1). With 85 participants, a maximum of 255 explanations were possible. However, a few participants left an explanation completely blank, and therefore we were not able to score three explanations, yielding a total of 252 scored explanations. The lead author and a science education expert independently scored participants’ explanations. Initial rater scores were at a high level of agreement ($r = .94, p < .01$). We discussed differences in our initial scores and reached full consensus on every explanation.

Figure 2 shows the frequency of scores for each evidence-to-model link. In the remainder of this section, we highlight the links in which participants wrote the greatest and second greatest number of explanations, as well as the link with the least number of explanations, to illustrate the distribution of scores. Participants wrote the greatest number of explanations ($n = 74; 29.4\%$ of the total explanations) for the link between Evidence #1 and Model A. Evidence #1 describes how atmospheric carbon dioxide concentrations have increased over time, and how carbon dioxide emissions due to human activities have also increased over time. Model A is the human-

| Table 2. Types of evaluation scoring rubric for explanatory tasks. |
| --- | --- | --- |
| Category | Description | Score |
| Erroneous evaluation | Explanation contains incorrect relationships between evidence and model, excluding misinterpreting a ‘Nothing To Do With’ relationship by elimination-based logic. The explanation may also be mostly inconsistent with scientific understanding and/or include nonsensical statements. | 1 |
| Descriptive evaluation | Explanation contains a correct relationship without elaboration, or correctly interprets evidence without stating a relationship. For example, the evidence-to-model link weight states that the evidence has nothing to do with the model. Explanation does not clearly distinguish between lines of evidence and explanatory models. Explanations could also demonstrate ‘elimination-based logic’ to come to a positive or negative weight, when evidence-to-model link weight states that the evidence has nothing to do with the model. For example, an explanation states that evidence supports one model, but the explanation also says that the evidence contradicts the other model. | 2 |
| Relational evaluation | The explanation addresses text similarities, and includes both specific evidence and an associated model or reference to a model. For example, explanation is correct, with an evidence-to-model link weight of strongly supports, supports, or contradicts as appropriate. Explanation distinguishes between lines of evidence and explanatory models, but does so in a merely associative or correlation manner that is often based on text similarity. | 3 |
| Critical evaluation | Explanation describes a causal relationships and/or meaning of a specific relationship between evidence and model. For example, explanation is correct, with an evidence-to-model link weight of strongly supports, supports, or contradicts as appropriate and reflects deeper cognitive processing that elaborates on an evaluation of evidence and model. Explanation distinguishes between lines of evidence and explanatory models, allows for more sophisticated connections, and/or concurrently examines alternative models. | 4 |
induced model of climate change. For this evidence-to-model link, the predominant score was 3 (relational evaluation; \( n = 62 \)). This indicates that most of the participants’ explanations discussed only the similarity in wording between Evidence #1 and Model A.

Participants wrote the second greatest number of evidence-to-model link explanations (\( n = 45; 17.9\% \) of the total explanations) for Evidence #2 and Model B. Evidence #2 describes the association between energy output by the Sun and average global temperatures over the past 100 years. Model B attributes current climate change to increasing amounts of energy released from the Sun. Interestingly, the predominant score for this link was 1 (erroneous evaluation; \( n = 30 \)) because many participants wrote that Evidence #2, which reveals that solar activity has been decreasing, supported Model B, which states that our current climate change is caused by increasing amounts of energy released from the Sun. In fact, a correct explanation of this link should say that Evidence #2 contradicts Model B. The fewest explanations (\( n = 12; 4.8\% \) of the total explanations) were written for the link between Evidence #4 and Model A. Evidence #4 describes paleoclimatic associations between sunspots (one measure of the Sun’s activity) and average global temperatures as measured by tree rings.

**Associations**

Most of the associations between types of evaluation for each evidence-to-model link, plausibility perceptions about models explaining the causes of current climate change, and knowledge about human-induced climate change were not significant. However, participants’ evaluation of the link between Evidence #2 and Model B and knowledge about human-induced climate change was significantly and moderately correlated (\( n = 45, r = .36, p = .015 \)). This link is the only one that shows evidence contradicting a model. Participants’ evaluations were dichotomous, in that some (\( n = 15 \)) demonstrated both relational and critical evaluation in their explanations and the remainder (\( n = 30 \)) evaluated this link erroneously.

Figure 2. Frequency of scores for each evidence-to-model link. In the figure, evidence-to-model links are coded based on the evidence number (1, 2, 3, or 4) and model (A or B) at each end of the link (e.g. E1_MA therefore shows the link from Evidence #1 to Model A).
Predictors of postinstructional knowledge

Because we found a significant association with evaluations of the Evidence #2 to Model B link with postinstructional knowledge, we conducted a multiple regression analysis to examine how well these evaluations, as well as model plausibility perceptions, predicted knowledge. Prior to conducting the regression, we screened the results and found that our analysis met the normality and linearity inherent in linear regression designs. There were no multivariate outliers (as measured by Mahalanobis distance) or univariate outliers (as measured by z-values). For the regression model, $R^2$ was significant, $F(2,44) = 7.1, p = .002$, with about 25% of the variance in postinstructional knowledge explained by the independent variables (types of evaluation of the Evidence #2 to Model B link and model plausibility ratings). Both evaluation category and plausibility were significant predictors of knowledge ($\beta = .32, p = .019$ for evaluation; $\beta = .35, p = .013$ for plausibility), where more sophisticated evaluations and higher plausibility perceptions predicted greater knowledge about human-induced climate change.

Plausibility emerging as a significant predictor in the regression is noteworthy because in and of itself, plausibility was not significantly associated with postinstructional knowledge for this evidence-to-model-link. This regression model therefore provides some preliminary evidence supporting the dynamic relationship between critical evaluation and plausibility perceptions, as suggested by Lombardi et al. (2016). Furthermore, this dynamic relationship between evaluation and plausibility may be of practical significance to both educational researchers and science educators as indicated by the relatively large effect size on knowledge scores (i.e. with $R^2 = .25$). We stress that these scores reflect general knowledge of climate change, not just knowledge about the information related in the specific evidence to model link. This large effect therefore indicates a potentially synergistic effect of critical evaluation, plausibility, and deeper understanding.

Discussion and implications

Four types of participants’ evaluations of evidence-to-model links emerged from the qualitative analysis: (a) erroneous evaluations that were inconsistent with scientific understanding, (b) descriptive evaluations that only superficially distinguished between lines of evidence and explanatory models, (c) relational evaluations that indicated a greater elaborative commitment but still made judgments based in similarity of evidence and model text, and finally, (d) critical evaluations, where causal relationships between evidence and the alternative models showed the greatest degree of elaboration and reasoning. These evaluation types reflected both students’ reasoning about scientific topics (Driver et al., 1996) and the spectrum of engagement needed for knowledge construction and reconstruction (Dole & Sinatra, 1998).

The categories revealed by our qualitative analysis may be a useful tool for both science educators and researchers to help gauge students’ types of evaluation as they consider alternative explanations of scientific phenomena. Specifically, such a tool may be useful in facilitating argumentation in the science classroom by developing students’ understanding and use of epistemic criteria to construct scientifically valid knowledge. In fact, educators and researchers may find such a tool valuable because of increased on having students evaluate scientific explanations. For example, one performance expectation in
newly developed *Next Generation Science Standards* says that students should ‘Evaluate the claims, evidence, and reasoning behind the idea that electromagnetic radiation can be described either by a wave model or a particle model, and that for some situations one model is more useful than the other’ (NGSS Lead States, 2013, p. 100). In Finland, the practice of evaluation is embedded within student problem solving groups that permeate the entire curriculum (Darling-Hammond, 2010). Under such an instructional setting, the rubric developed in this study might serve as one resource for developing and implementing a problem solving science classroom.

The types of evaluation described here relate specifically to the MEL diagram activity, which are efficient replacements for instructional materials that merely provide information. Teachers can easily employ MEL diagrams in one class meeting and immediately begin building a scientific habit of evaluation in students. However, despite the present study’s focus on the MEL, the results of our qualitative analysis can be easily modified for other instructional strategies used in science classrooms, such as use of critical questions and argument vee diagrams (Nussbaum & Edwards, 2011), metacognitive prompts (Peters & Kitsantas, 2010), openness to alternatives (Meyer & Lederman, 2013), peer-evaluation of constructed explanations (Wang, 2015), and self-regulation checklists (Peters, 2012).

The quantitative results revealed an interesting relationship between students’ evaluations, their perceptions of plausibility about the causes of current climate change, and knowledge of human-induced climate change. Specifically, evaluations of the link between contradictory evidence and models explaining the cause of current climate change, as well as model plausibility ratings, were significantly associated with postinstructional knowledge. These results suggest that science educators may wish to stress the importance of contradictory evidence in changing plausibility judgments about explanatory models, not only to demonstrate the process of scientific evaluation, but also to deepen student understanding of scientific content (Erduran & Dagher, 2014). As Bachelard (1968) states, ‘Two people must first contradict each other if they really wish to understand each other’ (p. 114). Students should also engage in explicit and conscious plausibility reappraisal of alternative theories to increase their knowledge about the nature of science. Students should understand that scientific explanations – such as theories and hypotheses – are tentative (Lederman, 1999). But more importantly, students should know ‘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,’ and ultimately ‘that predictions or explanations can be revised on the basis of seeing new evidence or of developing a new model that accounts for the existing evidence better than previous models did’ (National Research Council, 2012, p. 251). Therefore, engaging in critical evaluation to reappraise plausibility is a skill that may facilitate students’ development of the ability to reason scientifically and understand how scientific knowledge is constructed via collaborative argumentation.

Participants demonstrated various types of evaluation, with critical evaluation demonstrated at a relatively low frequency. However, we stress at this point that the students participated in the MEL activity for only about 90 minutes of instructional time. As we discussed earlier, Lombardi et al. (2013) showed that students experienced both statistically and practically significant shifts in plausibility toward the scientific model of human-induced climate change, as well as increases in knowledge that were sustained
six months after engaging in the activity. Therefore, we hypothesize that repeated use of instructional scaffolds such as the MEL may increase students’ ability to be critically evaluative when they are confronted with competing explanations of various scientific phenomena. Such scaffolds may be particularly relevant for controversial and abstract topics where a large gap exists between students’ and scientists’ plausibility perceptions (e.g., conversion of wetlands by land development, hydraulic fracturing in fossil fuel drilling). Repeated engagement with alternative explanations of such topics could help develop students’ scientific thinking because a ‘key activity of scientists is evaluating which … alternative does, or does not, fit with available evidence and, hence, which presents the most convincing explanation for [a] particular phenomenon’ (Osborne, 2012, p. 936).

Participants in the present study completed the MEL diagram and explanatory tasks individually (i.e., without interacting with other students and the teacher). Future studies could expand on the present study by looking at the effects of collaborative argumentation on interpretation of evidence-to-model links and construction of explanations, as well as the influences on the degree of evaluation expressed by the collaborative group. On the one hand, collaborative argumentation could promote greater elaboration and error detection through discourse (Nussbaum, 2011). On the other hand, students may experience lower levels of elaboration due to the influence of peers and the negative effects of motivated reasoning (Kunda, 1990; Taber & Lodge, 2006). The influence of peers on critical evaluation and plausibility reappraisal may be especially relevant for socio-scientific issues because of shared cultural values (Bencze, Sperling, & Carter, 2012; Jiménez-Aleixandre & Erduran, 2007).

Limitations and conclusions

The study participants were a representative sample of middle school students in many urban southwestern US school districts (i.e., predominantly Hispanic, with a relatively high proportion – just about half – in a low socioeconomic status). However, individuals should exercise caution generalizing these results beyond this population. Furthermore, this raises the need for future work on the relationships between critical evaluation, plausibility reappraisal, and knowledge construction with various other populations and age levels.

We also acknowledge that our results should be cautiously viewed because only one evidence-to-model link was significantly related to postinstructional knowledge. However, we are encouraged because this relationship was associated with a line of contradictory evidence. This suggests to us that additional research is warranted about the influence of contradictory evidence on both plausibility reappraisal and knowledge construction. With evaluation being placed as a pivotal scientific practice in which students should engage (National Research Council, 2012), science education researchers should endeavor to better understand how to help students evaluate levels of agreement and disagreement between evidence and alternative explanations. Giere, Bickle, and Mauldin (2006) say that agreement between evidence and explanatory models ‘may be a matter of degree’ (p. 31). Therefore, evaluation of this connection may be optimized when judging the fit between lines of evidence and an explanation, while simultaneously considering the fit with at least one other alternative explanation. In other words, equipping students with the evaluative
tools necessary to determine the best of all plausible alternatives is important to help them achieve, in part, the goals of current science education reform efforts to both deepen their understanding of scientific content and engage in scientific practices. Instruction that develops fundamental thinking skills (e.g. critical evaluation and plausibility reappraisal) is essential to developing a society that characteristically exhibits scientific habits of mind and is equipped to constructively deal with local and global challenges.

Note

1. The mode and structure of the model-evidence link (MEL) diagram was originally developed by a team of researchers at Rutgers University under the NSF-supported Promoting Reasoning and Conceptual Change in Science project for use in middle school life science classrooms (Chinn & Buckland, 2012). Lombardi et al. (2013) developed the climate change MEL used in the present study.

Acknowledgments

Part of this analysis emerged from the first author’s doctoral dissertation while at the University of Nevada, Las Vegas, with special thanks to the dissertation committee members, Dr Gale M. Sinatra, University of Southern California, and Dr E. Michael Nussbaum, Dr Gregory Schraw and Dr Matthew Lachniet, the University of Nevada, Las Vegas. We would also like to thank Dr. Janelle M. Bailey, Temple University, for her help in scoring students’ explanations.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Part of the analysis was also supported by the US National Science Foundation (NSF) [grant number DRL-1316057]. Any opinions, findings, conclusions, or recommendations expressed are those of the authors and do not necessarily reflect the NSF’s views.

Notes on contributors

Doug Lombardi is an Assistant Professor of Science Education in Temple University’s College of Education. Dr. Lombardi earned his PhD in Educational Psychology from the University of Nevada, Las Vegas, USA, a MS in Education from the University of Tennessee, Knoxville, USA, a MS in Environmental Engineering from the University of Tennessee, Knoxville, USA, and a BS in Mechanical Engineering from the University of Colorado, Boulder, USA. He is a licensed physical science, physics, and mathematics teacher, with a variety of classroom, professional development, and education and public outreach experience. His research is on the role of reasoning in conceptual change, epistemic cognition, and science learning.

Carol B. Brandt is an Associate Professor of Science Education and earned her PhD in Educational Thought and Sociocultural Studies from the University of New Mexico. She also holds a MS in Botany from Colorado State University and BA in Anthropology from Northwestern University. She completed a two-year postdoctoral research fellowship at the Center for Informal Learning and Schools at the University of California, Santa Cruz. Her research explores the anthropological and sociocultural dimensions of learning about science and the environment beyond the classroom and the ways that language structures participation as youth and adults move between home,
community, and school. Her ethnographic research examines the ways in which participants construct knowledge about the environment through citizen science projects and after school engineering programs.

**Elliot S. Bickel** is an undergraduate student at Temple University and is currently studying Mathematics with Teaching through Temple’s TUteach program. Elliot plans to graduate with a BS in Mathematics, as well as secondary teaching certification for mathematics and physics, in 2017. Elliot works as an undergraduate research assistant on a project funded by the National Science Foundation. He specifically analyzes students’ explanations of Earth science topics.

**Colin Burg** earned his BS in Middle Grades Education from Temple University.

**References**


Appendix

Excerpt on falsifiability from the pre-activity titled, ‘How do scientists change their plausibility judgments?’: Scientific ideas must be falsifiable. In other words, scientific ideas can never be proven. But, ideas can be disproven by opposing evidence. When this happens, scientists must revise the idea or come up with another explanation. Falsifiability is a very important principle when evaluating scientific knowledge.