

Student Decision Making in a Scenario-based Investigation of an Ecosystem

Cathy Ellen Mehl ^{1*}, Hui Jin ², Kenneth F. Llort ²

¹ Ohio State University, USA

² Educational Testing Service, USA

Received 28 June 2019 • Revised 3 August 2019 • Accepted 11 September 2019

ABSTRACT

Argumentation is an important component of scientific education (Osborne, 2010). However, how students create and evaluate competing arguments in scientific investigations is a complex construct, which presents significant challenges for assessment. We engaged 349 middle and high school students in a virtual scientific investigation based on an authentic problem involving the ecosystem of Yellowstone National Park. Students evaluated three possible causes for vegetation loss in Lamar Valley: tourism, global warming, and killing wolves. Students compared their responses to individual claims to formulate a final decision. The authors developed a coding system to assess student arguments and analysed how students changed their evaluations across the investigation. Four distinct levels of responses reveal the range of student decision making and patterns that have implications for instruction and assessment. Few student evaluations explained both merits and weaknesses of claims.

Keywords: argumentation, assessment, decision making, ecosystems, scientific investigation

INTRODUCTION

Clark and Sampson (2007) have defined argumentation as “a process of proposing, supporting, evaluating, and refining ideas in an effort to make sense of a complex or ill-defined problem” (p 296). Argumentation, as a core epistemic practice in science, has been advocated as an essential goal for science education (Bricker & Bell, 2008; Duschl, Schweingruber, & Shouse, 2007). In *A Framework for K-12 Science Education* (NRC, 2012), which informed the development of the Next Generation Science Standards (NGSS Lead States, 2013), argumentation is identified as one of the eight essential science practices that students should engage in to learn science content. There is consensus that experience with argumentation assists students in understanding the practice of scientists (Driver, Newton, & Osborne, 2000; Osborne, 2010). Empirical evidence also suggests that engaging in argumentation enhances conceptual understanding (Mercer, Dawes, Wegerif, & Sams, 2004; Venville & Dawson, 2010; Zohar & Nemet, 2002) and promotes critical thinking (Cavagnetto & Hand, 2012; Erduran & Aleixandre, 2008). As such, the ability to construct and evaluate arguments is a key component of scientific literacy (Cavagnetto, 2010; DeBoer, 2000; Erduran & Aleixandre, 2008; Khishfe, 2013). Following other science education researchers (Clark & Sampson, 2008; Osborne, Erduran, & Simon, 2004), we use the term, argument, to refer to the structure that includes a claim, evidence, and reasoning. More specifically, claim is a conclusion that conveys a point of view. Evidence refers to the data or facts that support the claim. Reasoning explains how evidence supports the claim.

Scientific argumentation differs significantly from traditional classroom activities because it requires students to evaluate evidence in ill-defined or complex problems instead of confirming a theory or illustrating a concept (Osborne, 2010). The existing research has well documented students’ learning difficulties with scientific argumentation. Students often have limited experience evaluating and justifying explanations (Kelley & Takao, 2002; Krajcik et al., 1998; Sadler, 2011; Sandoval & Reiser, 2004). Although many students are proficient in advancing claims, they find it challenging in evaluating evidence and providing reasoning or rebuttals (Deng & Wang, 2017; Novak & Treagust, 2017). Students often fail to include sufficient evidence and evidence based upon

Contribution of this paper to the literature

- The study models the use of a scenario-based task to engage students in an authentic investigation of an important and complex ecosystem concept: trophic cascade.
- The coding system developed assesses four levels of student decision-making among competing claims and captures whether student arguments are one-sided or two-sided.
- The authors propose a method for examining how students change their claims as new evidence emerges over the course of a virtual investigation.

appropriate science content (Clark & Sampson, 2008), struggle to provide multiple justifications (McNeill, Lizotte, Krajcik, & Marx, 2006; Sandoval & Millwood, 2005; Zohar & Nemet, 2002;), and have difficulty using the logical structure of formal arguments and applying science concepts (Erduran, Simon, & Osborne, 2004; Jin, Mehl, & Lan, 2015).

Although the above-mentioned studies have made important findings about students' learning difficulties, one area that has been largely under-researched is how students compare competing claims and select and support the most convincing claim. This topic has long been considered a cornerstone of argumentation. Kuhn (1992) suggested that "only by considering alternatives – by seeking to identify what is not – can one begin to achieve any certainty about what is (p. 622)." Sadler et al. (2004) pointed out that evaluating conflicting ideas in science prepares students for making everyday decisions about information presented in the news. Providing students with opportunities to evaluate evidence from competing claims to find the best explanation for a natural phenomenon is also called for in the Next Generation Science Standards (NGSS, 2013). Examining how students respond to competing claims can be challenging, since it requires comparison of multiple competing arguments and adjustment of claims with emergent evidence. Donnelly, Namdar, Vitale, Lai, and Linn (2016) investigated how seventh graders compared the views of evolution from Darwin and Lamarck in an argumentation writing task. Students in this study read about three scientific experiments in the context of evolution. Students were prompted to either elaborate on their ideas or to elaborate including ideas from both Lamarck's and Darwin's perspective. Both groups significantly grew in understanding of evolution concepts. The authors suggest that evaluating competing claims could promote deeper understanding than when students elaborate on ideas when guided to do so. Notably, students did not evaluate which claim was more reasonable or convincing. Novak and Treagust (2017) examined how students adjusted their claims in response to emerging data. In their study, fifty-eight seventh grade students completed an extended investigation of a complex phenomenon: water quality. The students wrote explanations of their claims four times over a six-week period as they collected data that sometimes supported original claims and sometimes did not support those claims. About 50% of the students were able to incorporate the new information and revise their claims. The authors identified several patterns in how remaining students dealt with emerging data, including ignoring the data that did not support initial claims, creating separate claims, and providing explanations with clear inconsistencies.

Another suggestion on promoting students' argumentation competency is engaging students in construct and evaluate arguments in an authentic context (Jiménez-Aleixandre, 2002; Sandoval & Millwood, 2005). Davenport, Milks, and Van Tassel (2015) point out that real world data are often imperfect, and that scientists seek the most convincing explanation rather than confirm one. They explain that evaluating authentic data challenges students to "grapple with complexity and draw conclusions in ways that parallel the work of scientists" (p 274). The recent studies took place in instructional contexts over time. The Donnelly et al. study (2015) required seven class periods while the Novak and Treagust study took place over 10 weeks. Assessment often occurs in one hour or class period. Our study adds to this important body of knowledge using an assessment tool that engages students in evaluating three competing claims using complex emerging data in an authentic scientific investigation.

In summary, relevant literature suggests that students should learn how to deal with competing claims and one effective way to promote students' argumentation practice is to engage them in constructing and evaluating arguments about issues in real-world situations. Therefore, we designed a scenario-based task to assess students' competency in dealing with competing claims in an authentic context: the loss of vegetation in the Lamar Valley of Yellowstone National Park. Ecosystem is a core content area that is addressed across the grade bands (NGSS Lead States, 2013), so students from both middle school and high school will have prior knowledge of the topic. In the investigation, students made hypotheses, changed their positions when new evidence emerged, and made conclusions by examining all existing arguments. As such, the scenario-based task offers insight into both how arguments are generated in the investigation and how students change their positions throughout the investigation. Our research question is: How do secondary students compare competing claims to select the most convincing claim? How do secondary students use evidence to support the selected claims? As an exploratory study, the purpose is to identify patterns of comparing, selecting, and supporting claims. Future research will validate those

patterns using more tasks and including more science topics, based on which, the comparison of student proficiency across school levels is meaningful.

METHOD

Students in this study evaluated three possible explanations for vegetation loss in the Lamar Valley of Yellowstone National Park, then made a decision about which claim was most convincing. A prior study focused on student responses to individual claims as they worked through the scenario-based task. This informed our approach to the current study, in which we focussed on how students conclude the investigation by reviewing their responses to individual claims and making a decision. We also examined how their final decision compared to choices they made during the investigation about the plausibility of individual claims.

Participants

The participants were 349 secondary students from four states: Pennsylvania, Georgia, Indiana, and Tennessee. These students included 199 high school students and 150 middle school students. Regarding race, 210 of the participants (60.2%) identified as Black or African American, 75 (21.5%) White, 54 students (15.5%) identified as Hispanic/Latino, 7 (2.0 %) identified as "other", and three students (0.9%) identified as Asian American. About 56% of our participants were female and 44% male. A large percentage (79.4%) of participating students received free or reduced-price lunch. A small number of students did not complete all the tasks or responded with random letters on at least one occasion. We did not include these responses in the assessment. As a result, we analysed responses from 320 students.


Instrument

The assessment tool that we used was a scenario-based task with questions that students completed on a computer during their regular course period. A scenario-based task is a series of questions linked to a common context; these questions collectively assess a complex construct (Bennett, 2010, 2011). The student is introduced to an imaginary team of students who present information about a particular phenomenon or problem. This information is analysed and used by the participant to form responses within a structure that scaffolds participation. This encourages depth of processing and assesses higher order thinking skills. It supports students through the process by including an advisor and three virtual students who provide their ideas, as well as evidence.


We developed a scenario-based task based on a real scientific investigation about the loss of vegetation in the Lamar Valley of Yellowstone National Park (Ripple & Beschta, 2003; Ripple, Larsen, Renkin, & Smith, 2001). In the early 1900s, vegetation in the valley was lush and many species coexisted. Predator control was widespread during this time, and by the 1940's wolf packs were rare. The population of elk increased at a dramatic rate in the years to follow and the landscape was barren by 1990. The relationship of predator removal on vegetation loss is an example of an important biological phenomenon called the "trophic cascade." Ripple et al. (2016) defined trophic cascade as "indirect species interactions that originate with predators and spread downward through food webs" (p. 846). The trophic cascade is crucial for evaluating the different claims about the cause of the vegetation loss but may be challenging to understand because it involves systems thinking (Jin, Shin, Hokayem, Qureshi, & Jenkins, 2019).

The task has three sections. In the first section, students observe pictures of a natural phenomenon to investigate—vegetation loss in Lamar Valley. The guiding question presented to students is, "What happened to all the plants?" [Figure 1](#) shows a screenshot of part of the introduction. In the second section, each of the three virtual students in the scenario-based task presents a possible explanation with some data about that claim. Student participants evaluate the evidence of each claim in turn for relevance and determine if the evidence supports or does not support the claim. [Figure 2](#) shows an example of how students view information about one claim. In the third section, students review the arguments they have constructed over the course of the investigation to generate a final decision about which claim is most likely to be true. They also provide their reasoning for this final decision in this third section of the task. [Figure 3](#) is an example of such a review generated from one student in the study.

ETS CBAL Question 7 of 16



A photograph of part of Lamar Valley from early 1920's



A Photograph of part of Lamar Valley from 1990's

Photo courtesy of commons.wikimedia.org/wiki/File%3ALamarRiver1969.jpg

Directions: Please compare the two photographs and answer the following questions.

How has the landscape of the Lamar Valley changed?


Do you think that the populations of animals in the Lamar Valley has changed?

Yes
 No

Please predict how the population sizes of animals have changed.

Figure 1. Introduction to the scenario

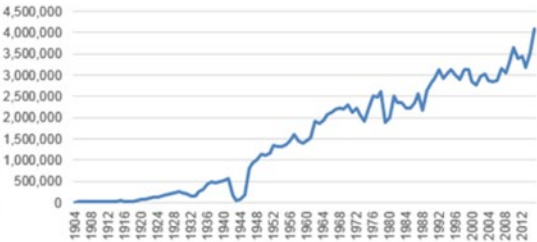
ETS CBAL Question 9 of 16



Sue

Here is some evidence I found about humans in the park, let's see if the evidence supports the claim that humans caused the plant loss.

Recreation Visitors to Yellowstone National Park



Guidelines for Viewing Wildlife (Cause 1)

- Use binoculars or telephoto lenses for safe viewing and to avoid disturbing them. By being sensitive to its needs, you will see more of an animal's natural behavior and activity.
- If you cause an animal to move, you are too close!
- It is illegal to willfully remain near or approach wildlife, including birds, within ANY distance that disturbs or displaces the animal.

Research Question: What caused the loss of plants in Lamar Valley?

Sue's Claim: I think tourism caused the loss of plants.

Do you find any evidence that is relevant to Sue's claim?

Yes
 No

Please identify all relevant evidence and explain whether the evidence is for or against Sue's claim.

Evidence	For	Against

Figure 2. Data about tourism

We now discuss the competing claims in detail. The first possible cause that students explore is tourism. In this task, students examine National Park Service data showing the number of visitors from 1900 to the present day, as well as park guidelines about safe viewing of wildlife. Students examine the evidence and select which evidence is relevant to the claim. They also identify whether the relevant evidence supports the claim or refutes it (i.e., "for" or "against"). Figure 2 illustrates the argument presented by the virtual student who claims that tourism is the cause of the loss of vegetation in Lamar Valley.

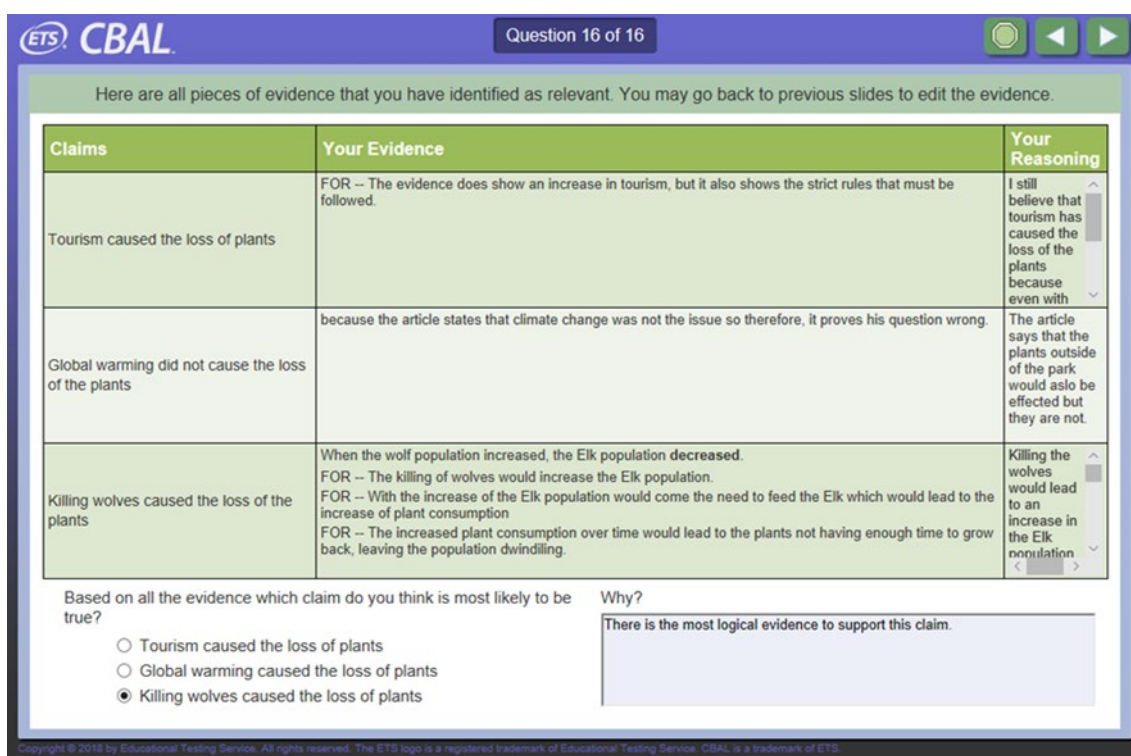


Figure 3. Summary of student data for final decision-making

The second possible explanation students examine is global warming. In this task, students review a scientific comparison of the plant populations inside and outside of the park. They also view reports from experts about the impact of climate change on plants. The information does not support the conclusion that global warming is likely to be the cause for the vegetation loss in Lamar Valley, thus the data are ‘anomalous.’ Students identify relevant evidence and note whether the evidence is ‘for’ the claim or ‘against’ it.

A virtual student named Christina provides the third task. Her evidence includes an article about the extirpation of wolves from the Yellowstone National Park. In the scenario, the three virtual students question the relationship between wolves and plants, at which point the virtual advisor provides more data about the wolf and elk populations following the reintroduction of wolves to the area in the 1990s. This data comes from a real life follow up study on wolf reintroduction (Ripple & Breschta, 2003). Students are once again asked to identify relevant evidence from this claim, and to determine whether each piece of evidence is “for” or “against” the global warming as a cause for the loss of vegetation.

Finally, the scenario-based program assists students in reviewing their responses for all three possible causes for the loss of the plants. These responses include evidence both “for” and evidence “against” a claim. They may go back to edit previous responses if they so choose. Students decide which claim was the most convincing and provide their reasoning. For their final decision, students may choose an explanation that they have previously supported or select an alternative they did not support during the investigation. They provide a justification for their final choice. Although students were asked to reflect on their previous responses to the scenario, many students used their own evidence at this point to support their choice of the most likely cause.

In Figure 3 we present an example of how previous responses would be made available to students, using mock data for illustration. Figure 3 demonstrates that a student may initially support more than one claim. The student would be asked to select only one claim from among the choices. The student could select one of the previously supported claims or chose to change to the previously unsupported claim. It is important to note that it was possible for students to reject all the claims during the investigation, yet they are required to make a final choice. This occurred in a small number of cases, and the authors agreed to consider these responses to be a ‘forced choice’ considering limited information. Analysing this information allowed the researchers to investigate how students’ ideas about competing claims change over the course of an investigation.

Data Analysis

The data analysed in this study includes the responses that students made to the two questions at the bottom of their summary page (see Figure 3). In the first question, they made a choice about which claim they found most

Table 1. A scale used to evaluate how students select and support the most convincing claim

Level	Description of the Level
1	The response does not evaluate evidence for the claim. It may be a simple restatement of the claim, a new claim, a fragmented sentence, or unintelligible. Example: "They gone."
2	The response is a general evaluation about the quality of the evidence or a value judgment about the claim. Example: "People don't care about the environment"
3	The response is an evaluation using logic and/or science, but with errors. The evidence used may be insufficient, inaccurate, or contain misconceptions. The link between the evidence and the claim may contain faulty logic. Example: "because when it gets hot the soil is dryer which means the water doesnt [sic] make it to the trees because it just get [sic] soaked up into the ground"
4	The response is an evaluation using sound logic and/or accurate science. There is sufficient and adequate use of evidence from the scenario; there is a logical link expressed between the evidence and the claim. Example: "Killing wolves brings back a lot of elk which they in turn eat more plants."

convincing: Tourism, Global warming, or Killing wolves. In the second question, students provided reasoning for their final decision. The following is a description of the analysis we performed on these extended responses in order to answer our research question.

We analysed students' responses in two steps. In the first step, we analysed a sample of 100 student responses to develop the coding rubric. We used grounded theory to uncover the patterns in the reasons students provided for their final choice (Glaser & Strauss, 1967). Based on this work, we developed an initial coding rubric. Next, in an iterative process of blind coding and comparison, we revised and refined this rubric. In this process, we discussed any discrepancies and collaboratively revised the rubric. The final rubric contains four distinct levels. Each level corresponds to a distinct approach toward the evidence in the scenario and reasoning the student uses to justify his or her position. Two researchers used the coding rubric to score all of the remaining responses independently. The agreement reached 90.4%. We discussed the responses coded differently in order to reach consensus on a final code.

FINDINGS

We evaluated the data from student final decisions in two ways. First, we assessed how the students constructed a final argument from their work on the three competing claims. Then we examined how final decisions compared with choices each student made on individual claims during the investigation. The following is a discussion of the results from these two evaluations.

Selecting and Supporting the Most Convincing Claim

As elaborated above, in the final decision page, students were provided with a summary of their previous responses to each possible claim. They were then asked to select the most convincing claim and support their decision with reasoning. Our research question examines how students select and support the most convincing claim. To answer this question, we first identified patterns in student responses. These patterns were used to develop a scale for evaluating student arguments about the claim that was the most likely cause for the loss of vegetation. This scale was revised through iterations of coding and discussion, and the rubric is presented in [Table 1](#). The scale contains four levels. Level 1 includes responses that do not provide information about the evidence or reasoning students may have used. Level 2 includes responses about the general value of the evidence or claim rather than logic or reasoning about specific evidence. Level 3 responses were more specific about evidence with associated reasoning but contained errors or flaws while Level 4 responses provided scientifically accurate evidence with logical reasoning.

As students examined the individual claims, they were prompted to consider whether each piece of evidence was "for" or "against" the claim. When students provided more specific evidence and reasoning in their final decision, many chose to mention only the evidence that supported the claim that they selected. Osborne et al. (2016) have coined the term "one-sided" to describe such arguments. More sophisticated and scientific arguments also address the weaknesses of other claims. Osborne et al. called these arguments "two-sided." We use the same distinction to further classify Level 3 and Level 4. Since the two-sided arguments are about the reasoning component, it is not relevant to Level 1 and Level 2. [Table 1](#) describes the patterns identified in each level.

Level 1

The simplest pattern in our data included responses that were brief, vague, or simple restatements of a claim. Students were asked which claim they considered to be the most likely explanation, and then to provide their reason. Some students seemed to confuse the claim and justification (e.g., "Because its tourism."). Others did not

elaborate or provide enough information to delineate why they decided upon the chosen claim (e.g., "They gone."). The students who received a code of Level 1 did not provide an evaluation of evidence, and their responses did not add to our understanding of student thinking.

Level 2

The students in this group showed more sophistication, but group members' evaluations were general. These "Level 2" students shared value judgments about the quality of the evidence or the quality of a claim, but they did not evaluate the evidence. Some students identified that there was "more" or "better" evidence for a certain claim but did not elaborate. Two other interesting subgroups emerged that revealed strong value judgments. One set of students spoke strongly about the destructive nature of humans (e.g., "People don't care about the environment."). Another set of students expressed commitments to climate change (e.g., "even though the evidence wasn't in my favor i [sic] still feel a though the global warming and other things had a negative effect on the vegetation").

Level 3

The largest group of students attempted to provide logic and science content in their evaluations, but their responses contained flaws. Three common patterns emerged within the data. First, some responses had insufficient evidence because students left out or ignored important details from the scenario. For example, one student reported, "The herbivores ate the plants." Although this is accurate, it does not fully explain the magnitude of the loss of vegetation in Lamar Valley compared with adjacent areas of the park. Second, some students provided inaccurate evidence. For example, a student suggested, "It is mostly that tourism had alot [sic] to do with it because the people were doing things that they shouldn't have done which cause the plant to go down." The scenario did not suggest this was happening, but it was a common theme. Third, some responses were not logically tied to the claim. For example, a student wrote, "killing wolves would increase the plants because wolves haven't been eating them if they are dead." It is unclear if the student thought that wolves ate plants directly or if the student failed to mention the impact of herbivores, such as elk. Thus, the logical connection has not been established.

These three patterns were sometimes observed in combination and could be further subdivided into two categories depending upon whether students addressed only the claim they were supporting, or also addressed weaknesses of other claims. In the following example, the student exhibits reasoning, but does not consider a relevant piece of evidence from the scenario- this kind of vegetation is growing normally outside the park. "Some of these animals wouldn't have food so nature [sic] die because they wouldn't have food to survive cause when it gets hot the soil is dryer which means the water doesnt [sic] make it to the trees because it just get soaked up into the ground." The argument is also one-sided in that it only addresses the claim that the student has chosen as most likely to be the cause for vegetation loss.

Two-sided arguments consider at least one claim that was not chosen as well as the one considered most likely to be true. In the example below, the student identifies what he or she sees as the weaknesses of the killing wolves and global warming claims but does not provide sufficient evidence to support the claim they have chosen most likely. Specifically, the student fails to consider the predator-prey relationship and its impact on local vegetation. "Because its the only reasonable answer. Wolves eat meat not plants and an article even proved that global warming wasn't happening and Yellowstone would not face global warming for years."

Level 4

The most sophisticated student responses provided sufficient and scientifically accepted evidence, connected by logical reasoning. These are Level 4 responses. In the following example, the student appropriately identifies the impact of declining wolf populations on elk, and the resulting effect on plants. He or she recognizes the mechanism of the trophic cascade as the reason for this impact, even though they have not used the scientific term. The argument is one-sided since it only addresses the claim most likely to be the cause. The student does not identify strengths or weaknesses of any other claim. "The killing of wolves decreased the wolf population. Then the elk population increased and to sustain them they had to eat more. Becasue [sic] they had to eat more, more vegetation [sic] was destroyed. And without a natural predator to decrease the elk population they continued to destroy the ecosystem."

In comparison, the student in the example below has discusses the weaknesses of the claims they have rejected, as well as the reason they have chosen the Killing Wolves as the most likely explanation. This two-sided argument addresses the impact of the wolves on elk, but also identifies the weaknesses of the global warming and tourism claims. "That would cause the biggest impact. Studies show that global warming wouldn't have an impact until the future, and tourists couldn't get close enough to the wildlife to have an impact. Killing wolves would allow [sic] the elk population to grow, making the plant population decrease."

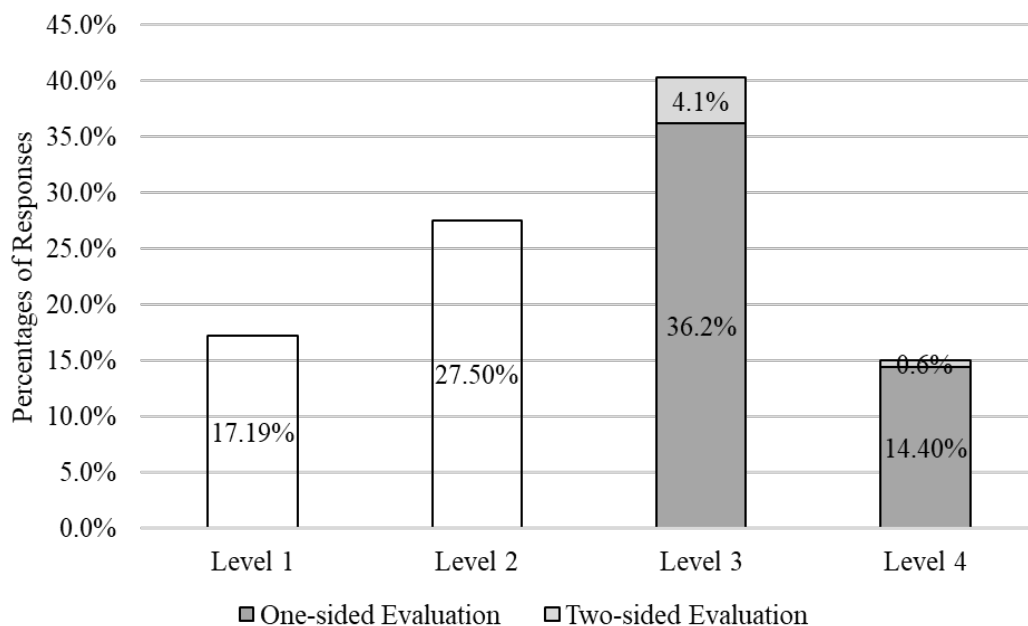


Figure 4. Percentage of students at each level by type of evaluation

Figure 4 presents the distribution of students' responses at different achievement levels. The great majority of responses were coded as either level 2 or level 3, indicating that students may need additional support to use scientific evidence and logic when making a decision about which claim is most likely. This is surprising, given the degree of scaffolding built into the scenario-based tasks. For example, scenario prompted students to describe evidence that was relevant in examining each claim, yet many students did not include such evidence in the reason for their final choice. One possible explanation for this finding is that students may think relevant means supportive, even though scientists value evidence that disconfirms a claim as well. When a student makes a commitment to an idea or claim that lacks supporting evidence within the task, they face a dilemma. They can bring in outside evidence that may or may not be scientifically accurate, or they can limit their reasoning to a general evaluation. The work of Novak and Treagust (2017) suggests that students who do not revise their claims when new evidence emerges either ignore the dissonant information or settle for inconsistencies. In our study, those who ignored important information would not have met the rubric for Level 4 based on inaccurate or incomplete evidence while those with inconsistencies were not logical in their reasoning.

One-sided versus two-sided evaluations

We asked students to select which claim was most likely, based on the evidence, and why. Previously, students had determined whether each piece of evidence was "for" or "against" an individual claim. Many students provided reasoning that reasoned about the evidence that supported the claim they selected. Osborne et al. (2016) use the terms "one-sided" and "two-sided" to clarify this qualitative difference in their learning progression for argument in science. A one-sided evaluative judgment occurs when a student compares two competing arguments and states why one claim is the most compelling without providing a warrant about the other claim. Two-sided evaluations explicitly state the value of the chosen claim and tell why the other claim was not as strong. Level 1 and Level 2 responses in our study were not explicit evaluations about the evidence, so such a classification did not apply.

Our study had three competing claims. We coded responses that only provided support for the selected claim as "one-sided" and coded those that addressed the weaknesses of another claim or claims as "two-sided" evaluations. Among Level 3 responses, only ten percent addressed weaknesses of a claim not selected as well as the reasons for choosing the most compelling claim. Only four percent of Level 4 responses provided information about the weaknesses of a claim or claims not chosen. An examination of the content of student responses indicate evidence 'for' a claim was much more frequent than evidence 'against' a claim. This is consistent with the findings of Osborne et al. (2016) and may represent the everyday notion that data must agree with a conclusion in order to be relevant. The nature of science values and even embraces evidence that disclaims a hypothesis, as well as evidence that supports a hypothesis. This is not a feature of everyday arguments. Encouraging students to seek evidence that disconfirms requires investigations with more complex, messy data (Davenport et al., 2015).

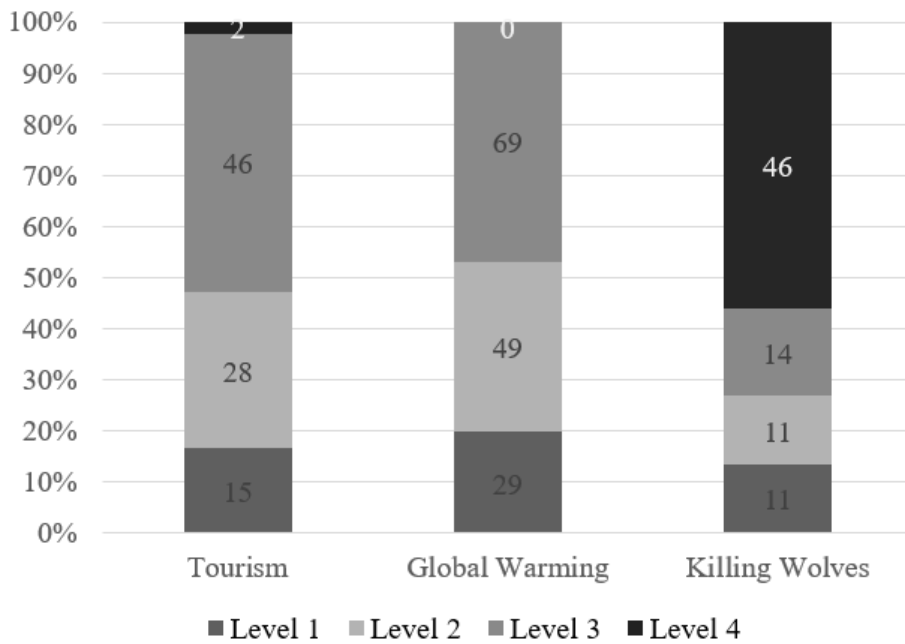


Figure 5. Number of students by level of response (N = 320)

When the level of student responses was analysed according to their final decision, the data revealed three important trends (see Figure 5). First, the majority of students selected global warming as the most likely claim. This was despite the fact that none of the evidence provided in the scenario supported the conclusion that global warming was causing vegetation loss. That is, there were no supporting data within the investigation, yet most students felt strongly about this explanation.

Second, a large number of students supported the idea that Tourism was the cause for the loss of vegetation. Two percent of their responses were coded Level 4 based on the logic of the explanation, but most of the responses relied on inaccurate or insufficient data. Students suggested rather dramatic explanations at times, in order to justify their decision (e.g., humans cut the trees down or ignored the rules). While most of the students who selected Global Warming or Tourism provided limited or general explanations, more than half of the students who selected Killing Wolves were able to provide logical and scientifically accurate reasons. In part, this reflects the level of scientific evidence available to support this claim, but it also means that students were able to connect this evidence to the claim in a logical manner. Many of the students who rejected this claim expressed the idea that wolves did not eat plants and were therefore not relevant. The mechanism of trophic cascade plays a critical role in decisions about this scenario. According to a learning progression understanding of ecosystems developed by Author et al. (2019), constructing a causal mechanism to explain a phenomenon is the highest level of achievement. It may be that many of the students in this study had not yet developed systems thinking.

Although ecosystem is a core content area that is addressed across the grade bands (NGSS Lead States, 2013), student argumentation skill may have been age dependent. We tested this by examining the results from students without missing data (i.e., those who had responded to each of the scenario tasks and provided reasons for their final decision). The number of middle school students who were rated at each level according to our scale was compared to the number of high school students at each level. Our results are reported as percentages in Figure 6. These findings suggest that middle school students were not less capable of forming higher level arguments about which claim was most likely to have caused the loss of vegetation in Lamar Valley. Middle school students were more likely than high school students to respond by providing evidence from the scenarios, and to provide more elaborate reasoning. It is unclear why this surprising result occurred, and is an important area for further research.

Evaluating and Comparing Competing Claims

We compared student choices within the investigation were to their final decision. Four patterns emerged. As described above, some students supported multiple possible claims during the investigation, and decided to 'PICK' from those claims the explanation they thought most likely. Other students only supported one claim and decided to 'KEEP' that claim as a final decision. Although it was rare, there were students who decided to 'CHANGE' their responses in making a final decision. Lastly, some students did not support any of the claims during the investigation and thus, had to make a 'FORCED CHOICE' in order to complete the task. Figure 7 shows the relative numbers of students making each of these moves according to final decision.

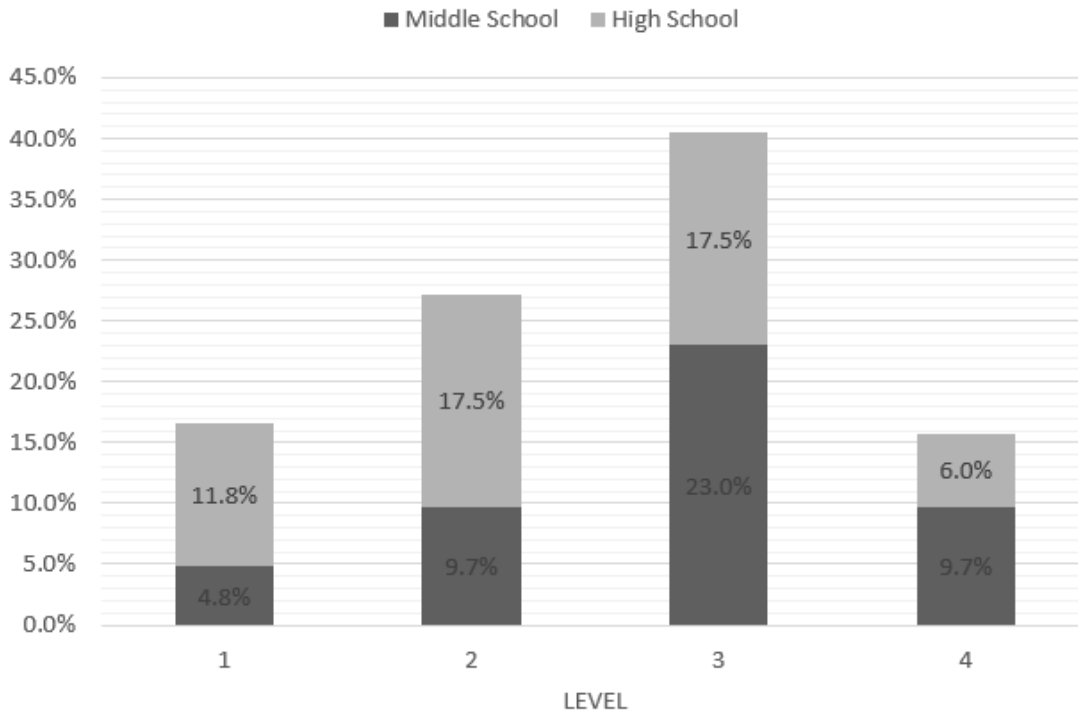


Figure 6. Percentage of students at each level by grade band (N = 331)

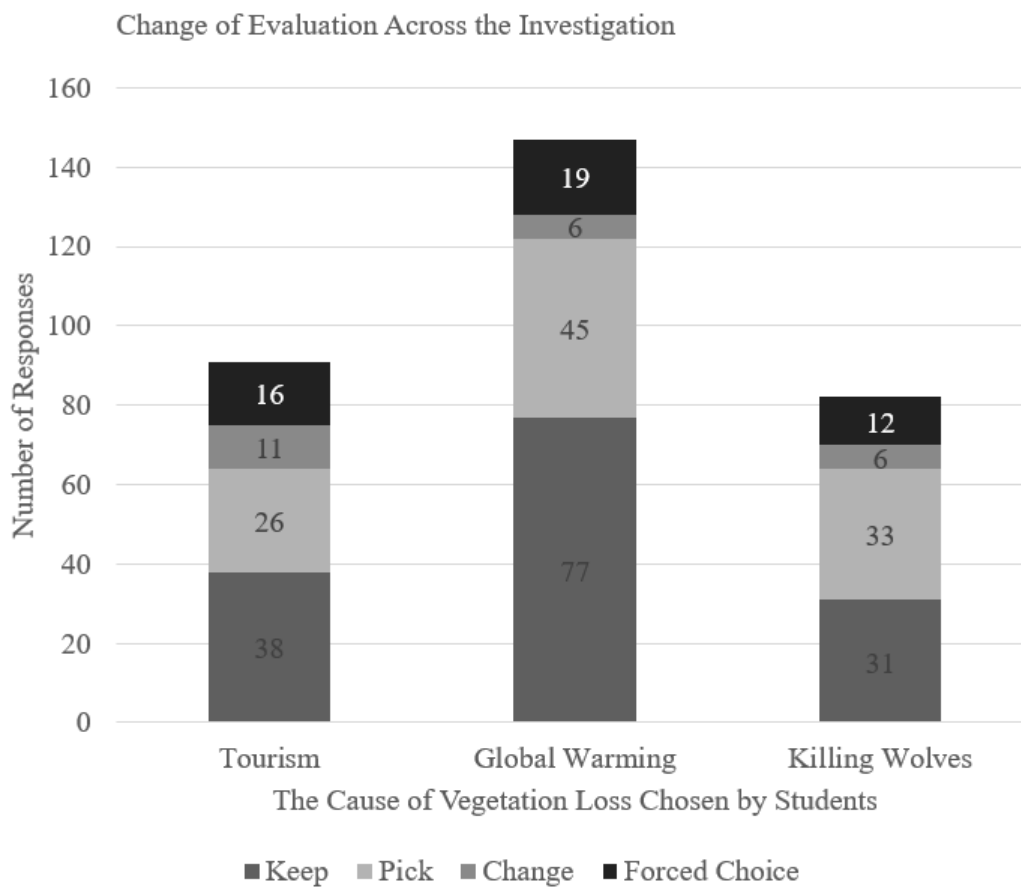


Figure 7. Summary of student data for final decision

A few surprising trends emerged from these data. First, the number of students who made a single choice, rejecting other claims and sticking to it, was overwhelming. If the original claim selected by a student turns out to be the most convincing, then there would be no reason to change. However, we observed this pattern for all competing claims. It suggests that the power of commitment to a cause is significant. Second, the smallest numbers of students were those who were willing to 'CHANGE' their support for a competing claim to an alternative claim. Although it is the nature of science to revise or reject a claim, it is not a common everyday practise. A final observation regards the number of students who made a 'FORCED CHOICE.' Given the amount of scaffolding and data support within the scenario-based task, it was surprising that so many students still found all of the explanations unsupportable initially. It would be of interest to find out more about this group.

CONCLUSIONS AND IMPLICATIONS

The computer-based scenario in this study provided many supports to students as they investigated a problem involving the ecosystem of Lamar Valley. The virtual students and an ecologist guided the examination of three possible explanations using authentic data. Students were prompted to review the evidence and provide reasoning to explain their choice. Despite the scaffolding present in the scenario-based task, many students gave only a claim or general statement of belief to justify their choice regarding what caused the loss of vegetation in Lamar Valley. One explanation is that students did not understand how to construct more sophisticated arguments. Scientific argumentation is a complicated skill. Perhaps more intervention is necessary for students to recognize structural features that are expected. Another possible explanation is that students' responses could have been more elaborate but chose not to provide further detail. This is consistent with the findings of Osborne and colleagues (Osborne, MacPherson, Patterson, & Szu; 2012). They note that students responding in written form to an audience they perceive as knowledgeable may not elaborate justifications, perhaps to avoid explaining to those they believe will already have the answers. Uncovering student motivations and thought processes behind this finding is an important area for further research. The scale we developed might be a useful assessment tool within such instructional contexts in which more qualitative follow up with students is possible.

Most students made final decisions that included an evaluation of evidence but fell short of logical and scientifically accurate reasoning (i.e., Level 3). In particular, the most popular explanation was about climate change despite anomalous data. The students may have drawn upon their own experience, perhaps revealing misconceptions or distortions to support a favoured claim. Even when scaffolding encouraged a re-evaluation of competing evidence, students primarily stuck with explanations they had initially supported. This is consistent with research, which indicates that students tend to limit their arguments to include only evidence that supports their claims while largely ignoring evidence that refutes their claims (Clark & Sampson, 2008; Kortland, 1996; Osborne et al., 2016; Sadler & Donnelly, 2006).

Another relevant finding involves the commitment many students expressed to beliefs about human nature or the environment. Everyday forms of persuasion involve this kind of effort to bring others over to one's way of thinking, so this finding should not be surprising. Recent research by McCrudden and Barnes (2016) regarding the 'myside bias' would also explain the patterns we observed. Students frequently referred to experience that confirmed the beliefs they held about the destructiveness of people or the impact of climate change. McCrudden and Barnes suggest that such beliefs shape the way that students evaluate evidence, even leading them to construct evidence that supports their ideas. One goal of science education is to enable students to act more like scientists in objectively examining the evidence, but complex contexts involving the environment are bound to be belief-laden. Providing students with more experience in analysing rich contexts with complex problems may change this learning problem.

Although it is important that students access their prior knowledge and values in an investigation, helping students move towards a more scientific treatment of competing explanations will involve a variety of supports. First, students need even more support to determine what counts as evidence in a scientific argument. This may result in fewer value laden judgments. Second, they need support to expand their use or inclusion of available evidence, considering both evidence for and against a claim rather than seeking confirmation only. Third, students may need specific support and practice evaluating both strengths and weaknesses of all competing arguments rather than focusing on the positive aspects of a selected choice. Finally, instructional settings may need to emphasize the refinement and revision of ideas more fully for students to embrace this scientific way of knowing.

At the conclusion of our investigations, we have uncovered some limitations to address in future studies. First, in many cases, the students were limited in the amount and type of evidence they included in the reason for their final decision. The current study was designed to be completed in one class period, so the amount of evidence presented to students was carefully chosen to provide discrete pieces of information for each competing claim while avoiding response fatigue. Students may have needed more support to determine what counts as evidence, or they may have needed more explicit expectations for the assessment task. Kelley and Takao (2002) identify six "epistemic levels" of evidence which range from the most specific observations (data) to more abstract models (interpretations

of graphs) and accepted scientific principles (trophic cascade). We believe the scenario in a follow up study could include a wider range of evidence for each competing claim in order to focus students on this important task. Secondly, the reading level of students may be a confounding variable not assessed within the current framework. If students experience difficulty reading the evidence, they may not include that evidence in their reasoning. Students at the lower grade levels may also be concrete thinkers. These students would look for direct relationships (e.g., wolves eat meat) and have more difficulty identifying the indirect impact of wolves on plants. Pretesting the new scenario with a limited number of representative students within an interview context might shed a great deal of light on these concerns.

REFERENCES

- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*, 70-91. <https://doi.org/10.1080/15366367.2010.508686>
- Bennett, R. E. (2011). CBAL: Results from piloting innovative K-12 assessments. *ETS Research Report, RR-11-23*. <https://doi.org/10.1002/j.2333-8504.2011.tb02259.x>
- Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education, 92*, 473-498. <https://doi.org/10.1002/sce.20278>
- Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K-12 science contexts. *Review of Educational Research, 80*, 336-371. <https://doi.org/10.3102/0034654310376953>
- Cavagnetto, A., & Hand, B. (2012). The importance of embedding argument within science classrooms. In M. S. Khine (Ed.), *Perspectives on scientific argumentation: Theory, practice and research* (39-53). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2470-9_3
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*, 293-321. <https://doi.org/10.1002/tea.20216>
- Davenport, K., Milks, J., & Van Tassell, R. (2015). Using evolutionary data in developing phylogenetic trees: A scaffold approach with authentic data. *The American Biology Teacher, 77*, 274- 283. <https://doi.org/10.1525/abt.2015.77.4.7>
- Deng, Y., & Wang, H. (2017). Research on evaluation of Chinese students' competence in written scientific argumentation in the context of chemistry. *Chemistry Education Research and Practice, 18*, 127-150. <https://doi.org/10.1039/C6RP00076B>
- Donnelly, D. F., Namdar, B., Vitale, J. M., Lai, K., & Linn, M. C. (2016). Enhancing student explanations of evolution: Comparing elaborating and competing theory prompts. *Journal of Research in Science Teaching, 53*, 1341-1363. <https://doi.org/10.1002/tea.21331>
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287-312. [https://doi.org/10.1002/\(SICI\)1098-237X\(200005\)84:3%3C287::AID-SCE1%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(200005)84:3%3C287::AID-SCE1%3E3.0.CO;2-A)
- Duschl, R. A., Schweingruber, H. A., Shouse, A. W., National Research Council (U.S.), National Research Council (U.S.), & National Research Council (U.S.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C: National Academies Press.
- Erduran, S., & Aleixandre, M. (2008). *Argumentation in science education: Perspectives from classroom-based research*. Dordrecht: Springer.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915-933. <https://doi.org/10.1002/sce.20012>
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. <https://doi.org/10.1097/00006199-196807000-00014>
- Jimenez-Aleixandre, M.-P., & Pereiro-Munoz, C. (2002). Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education, 24*, 1171-90. <https://doi.org/10.1080/09500690210134857>
- Jin, H., Mehl, C., & Lan, D. (2015). Developing an analytical framework for argumentation on energy consumption issues: Analytical framework for argumentation. *Journal of Research in Science Teaching, 52*(8), 1132-1162. <https://doi.org/10.1002/tea.21237>

- Jin, H., Shin, H., Hokayem, H., Qureshi, F., & Jenkins, T. (2019). Secondary students' understanding of ecosystems: A learning progression approach. *International Journal of Science and Mathematics Education*, 17(2), 217-235. <https://doi.org/10.1007/s10763-017-9864-9>
- Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86, 314-342. <https://doi.org/10.1002/sce.10024>
- Khishfe, R. (2014). Explicit nature of science and argumentation instruction in the context of socioscientific issues: An effect on student learning and transfer. *International Journal of Science Education*, 36, 974-1016. <https://doi.org/10.1080/09500693.2013.832004>
- Kortland, K. (1996). An STS case study about students' decision making on the waste issue. *Science Education*, 80, 673-89. [https://doi.org/10.1002/\(SICI\)1098-237X\(199611\)80:6%3C673::AID-SCE3%3E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1098-237X(199611)80:6%3C673::AID-SCE3%3E3.0.CO;2-G)
- Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, 7, 313-350. <https://doi.org/10.1080/10508406.1998.9672057>
- Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62, 155-172. <https://doi.org/10.17763/haer.62.2.9r424r0113t67011>
- 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. https://doi.org/10.1207/s15327809jls1502_1
- McCrudden, M., & Barnes, A. (2016). Differences in student reasoning about belief-relevant arguments: a mixed methods study. *Metacognition & Learning*, 11, 275-303. <https://doi.org/10.1007/s11409-015-9148-0>
- Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, 30, 359-377. <https://doi.org/10.1080/01411920410001689689>
- National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*.
- NGSS Consortium of Lead States. (2013). *Next generation science standards: For states, by states*. Washington D.C.: Achieve, Inc.
- Novak, A. M., & Treagust, D. F. (2018). Adjusting claims as new evidence emerges: Do students incorporate new evidence into their scientific explanations? *Journal of Research in Science Teaching*, 55, 526-549. <https://doi.org/10.1002/tea.21429>
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466. <https://doi.org/10.1126/science.1183944>
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41, 994-1020. <https://doi.org/10.1002/tea.20035>
- Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, 53, 821-846. <https://doi.org/10.1002/tea.21316>
- Osborne, J., MacPherson, A., Patterson, A., & Szu, E. (2012). Introduction. In M. S. Khine (Ed.), *Perspectives on scientific argumentation: Theory, practice and research* (3-15). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2470-9_1
- Ripple, W. J. & Beschta, R. L. (2003). Wolf reintroduction, predation risk, and cottonwood recovery in Yellowstone National Park. *Forest Ecology and Management*, 184, 299-313. [https://doi.org/10.1016/S0378-1127\(03\)00154-3](https://doi.org/10.1016/S0378-1127(03)00154-3)
- Ripple, W. J., Estes, J. A., Schmitz, O. J., Constant, V., Kaylor, M. J., Lenz, A., Motley, J. L., ... Wolf, C. (2016). What is a trophic cascade? *Trends in Ecology & Evolution*, 31, 842-849. <https://doi.org/10.1016/j.tree.2016.08.010>
- Ripple, W. J., Larsen, E.J., Renkin, R. A., & Smith, D. W. (2001). Trophic cascades among wolves, elk, and aspen on Yellowstone National Park's northern range. *Biological Conservation*, 102, 227-334. [https://doi.org/10.1016/S0006-3207\(01\)00107-0](https://doi.org/10.1016/S0006-3207(01)00107-0)
- Sadler, T. (2011). Teaching and learning about science: Language, theories, methods, history, traditions and value. (Book review). *Science Education*, 95, 186-187.
- Sadler, T., Chambers, W., & Zeidler, D. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*, 26, 387-409. <https://doi.org/10.1080/0950069032000119456>

- Sadler, T. D., & Donnelly, L. A. (2006). Socioscientific argumentation: The effects of content knowledge and morality. *International Journal of Science Education*, 28, 1463-1488. <https://doi.org/10.1080/09500690600708717>
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23, 23-55. https://doi.org/10.1207/s1532690xci2301_2
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 3, 345-372. <https://doi.org/10.1002/sce.10130>
- Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, 47, 952-977. <https://doi.org/10.1002/tea.20358>
- Walsh, E. M., & McGowan, V. C. (2017). 'Let your data tell a story:' climate change experts and students navigating disciplinary argumentation in the classroom. *International Journal of Science Education*, 39, 20-43. <https://doi.org/10.1080/09500693.2016.1264033>
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39, 35-62. <https://doi.org/10.1002/tea.10008>

<http://www.ejmste.com>