Articulating Uncertainty Attribution as Part of Critical Epistemic Practice of Scientific Argumentation

Hee-Sun Lee, Amy Pallant, Sarah Pryputniewicz, and Trudi Lord
hlee@concord.org, apallant@concord.org, spryputniewicz@concord.org, tlord@concord.org
The Concord Consortium

Ou Lydia Liu, Educational Testing Service, lliu@ets.org

Abstract: Models are important in discovering trends, developing and testing theories, and making predictions about complex systems. Since models cannot represent all known and unknown aspects of how nature operates, claims based on model-based data inevitably contain uncertainty. This study explores (1) how high school students attribute sources of uncertainty when prompted as part of an argumentation task and (2) how intelligent feedback may guide them to become more cognizant about deep uncertainty associated with model-based data. Phenomenological analyses of students' uncertainty attributions (N = 840) identified five distinct patterns: self-introspection, personal theories and experiences, data source acknowledgement, scientific description based on singular causal accounts and frequency of observations, and deep uncertainty based on epistemic or ontic accounts. Discourse captured on video illustrated how intelligent feedback enhanced uncertainty attribution.

Topic introduction

The Next Generation Science Standards (NGSS Lead States, 2013) encourage students to engage in practices through which scientific ideas are originated. Integrating scientific argumentation into science teaching is recommended because it allows students to participate authentically in sense making with data during investigation (Duschl & Osborne, 2002) and through communication (Kuhn, 2010). During investigation, students make claims based on data in light of their understanding of established knowledge (Bricker & Bell, 2008). In communication, students compare and contrast the strengths and weaknesses of evidence-based arguments (Erduran, Simon, & Osborne, 2004). Scientific argumentation as an epistemic practice generates two types of discourse: "theoretical discourse, pertaining to what theories of the world best fit the data, and practical, deliberative discourse, regarding how to apply those theories to reach practical goals" (Nussbaum, Sinatra, & Owens, 2012, p. 17). For example, students can use carbon dioxide data captured in ice cores as part of theoretical discourse concerning how changes in greenhouse gas concentrations contribute to atmospheric temperatures over time. Students also can engage in "deliberative" arguments about whether to impose carbon taxes based on various types of data. This study addresses the former, i.e., knowledge generation discourse rather than the latter dealing with arguments in sociocultural decision making.

Challenges to engaging students in argumentation are well documented in the literature: (1) students have difficulties in differentiating among claim, evidence, and reasoning (Berland & Reiser, 2009), (2) students lack experience in interpreting evidence in terms of theory (McNeill, Lizotte, Krajcik, & Marx, 2006), and (3) students lack epistemic commitment (Sandoval, 2003). In constructing a scientific argument, students need rhetorical support on how to write a convincing argument as well as content support on what knowledge should be used to interpret evidence. While scaffolding individual students is needed, it is unrealistic to expect a teacher to systematically intervene with every student in her class (McNeill & Pimentel, 2010). To address this issue, we developed an intelligent feedback system that (1) diagnoses students' arguments through an automated scoring engine based on machine learning algorithms developed for natural language processing and (2) provides students with immediate feedback matching their current performance.

This study addresses high school students' written as well as spoken discourse that occurred when they used a computer-based groundwater model to determine whether water that infiltrates is trapped underground. Evidence from the groundwater model for students to make claims was constrained due to the fact that groundwater systems in the real world cannot be modeled exactly. Students needed not only to select and use data from this imperfect model to make a knowledge-based claim, but also to recognize that their claim was constrained. This study explores two research questions: (1) how high school students attributed sources of uncertainty when prompted as part of model-based argumentation and (2) how intelligent feedback might guide students to become more cognizant about deep uncertainty associated with model-based data.

Theoretical framework

Argumentation is carried out through written or spoken discourse often accompanied by symbols, representations, and visualizations (Walton, Reed, & Macagno, 2008). The field-independent structure of arguments is well recognized (Toulmin, 1958) such as a *claim to answer* a driving question, *data* that support the claim, *warrants* that explain how data support the claim and how established scientific knowledge *backs* the warrants, *qualifiers* that indicate the strength of the claim given evidence and knowledge, and *conditions of rebuttal* where the claim may not be held true. Most research on written scientific argumentation has focused on characterizing and improving students' coordination between theory and evidence embodied in claim, data as evidence, and warrants and backing as knowledge-based reasoning (Sampson & Clark, 2008). Students' uses of qualifiers and conditions of rebuttal have mostly been studied as counterarguments or rebuttals in written argumentation (Erduran et al., 2004) or in social settings (Sampson & Clark, 2009).

However, qualifiers and conditions of rebuttal can play a different but still important role in written argumentation because they invoke the notion of uncertainty, i.e., the extent to which knowledge claims are bounded by evidence generated from particular investigation contexts. As Bogen and Woodward (1988) pointed out, scientific knowledge explains facts related to a phenomenon but not necessarily facts related to raw data that represent aspects of the phenomenon. Data to which students have access are "dependent upon the peculiarities of the particular experimental design, detection devices, or data-gathering procedures" (Boumans & Hon, 2014, p. 2) and are just one of many incidences that can exemplify some but not all aspects of the phenomenon. Drawing unwavering knowledge claims from data is almost impossible and thus involves a great degree of uncertainty. Staley (2014) characterized two modes of reasoning by scientists when working with data. In the *use* mode, scientists use theoretical and methodological assumptions to arrive at substantive conclusions from the data. In the *critical* mode, scientists carefully examine those assumptions. The current emphasis on promoting scientific argumentation in the classroom through claim-evidence-reasoning may not take full advantage of an instructional opportunity where students can also learn about the *critical* use of data.

Environmental science topics such as climate change have been commonly used in classrooms for scientific argumentation (Nussbaum, Sinatra, & Owens, 2012) based on publicly available data and simulation models by scientists (Spiegelhalter, Pearson, & Short, 2011). Since environmental systems are complex, there exists epistemic uncertainty due to fundamental limitations with investigators' theoretical and methodological abilities to understand and predict how nature works. There exists also ontic uncertainty because "the physical world has an element of irreducible elusiveness. The result of an experiment is not determined by the conditions under the control of the experimenter. The lack of control is not the experimenter's deficiency, but rather nature's indeterminism" (Ben-Haim, 2014, p. 165). As models are the main means of investigating and understanding environmental systems, both epistemic and ontic uncertainty sources associated with model-based data should be examined explicitly by a person who is presenting an argument.

Some uncertainty associated with scientific investigation or modeling can be quantified in probabilistic terms. However, there is *deep uncertainty* that "results from myriad factors both scientific and social, and consequently is difficult to accurately define and quantify" (Kandlikar, Risbey, & Dessai, 2005, p. 444). Kahneman and Tversky (1982) pointed out that uncertainty is part of everyday life because we act without full knowledge, information, or understanding of any encountered stimulus. They noted that uncertainty can be captured as confidence on "a prediction, estimate or inference to which one is already committed" (p. 150). Kahneman and Tversky (1982) also reported that people often attribute uncertainty in natural language to the external world to seek more objective criteria or to personal state of knowledge due to internal ignorance. Those who seek external attribution use either frequencies of occurrence across multiple similar cases (distributional) or causal propensities that explain a typical or exemplary case that proves their point (singular mechanistic). Internal attributions can be based either on personal theories and experiences irrespective of external criteria or on introspective confidence. We use Kahneman and Tversky (1982)'s framework as a starting point to categorize students' uncertainty attribution when they construct an argument.

Methods

Scientific argumentation task

The scientific argumentation task we analyzed in this study was embedded in an online curriculum module entitled "Will there be enough freshwater?" This water module consists of six activities that guide students to explore the distribution and use of fresh water on Earth. Students experiment with models to explore Earth's groundwater system. They are also introduced to scientific data about freshwater distribution and use on Earth. The module is designed for five 45-minute class periods. Throughout the module, students are engaged in the practice of scientific argumentation as they work with scientific data, observations, and computer-based simulations. There are eight argumentation tasks that are structured similarly to scaffold students: (1) multiple-choice claim, (2) open-ended explanation of claim using "Explain your answer," (3) five-point Likert scale uncertainty rating from not at all certain to very certain, and (4) open-ended explanation of uncertainty rating "Explain what influenced your certainty rating." Figure 1 illustrates the first argumentation task related to the movement of water in and out of the Earth's surface. This task structure was validated to measure uncertainty-infused scientific argumentation (Lee et al., 2014). The activity starts with the importance of the topic and asks students to tinker with the model to make observations about how precipitation moves through the various sediment layers in the ground with different degrees of permeability. Students take a snapshot of the model run and draw the longest path a water droplet may take. Students then answer a multiple-choice question asking which layer of the model blocks water from flowing through. When students choose an answer, the module immediately answers whether the answer is correct. These two questions are designed to help students interpret the data and the representation of the model and elicit the knowledge necessary for students to successfully complete the scientific argumentation task that follows. The driving question for the groundwater model reads, "When water is absorbed by the ground, is it trapped in the ground?" Students respond to the four argumentation prompts, which are placed within a blue border called "arg-block." For each of the four argumentation prompts, students can expand a set of hints. For instance, hints for uncertainty rating explanation are (1) A good certainty explanation will explain why you are certain or uncertain about your response; (2) Some topics are more certain than others; (3) Consider the completeness of the evidence, biases in the evidence, and changes that could affect the trends over time.

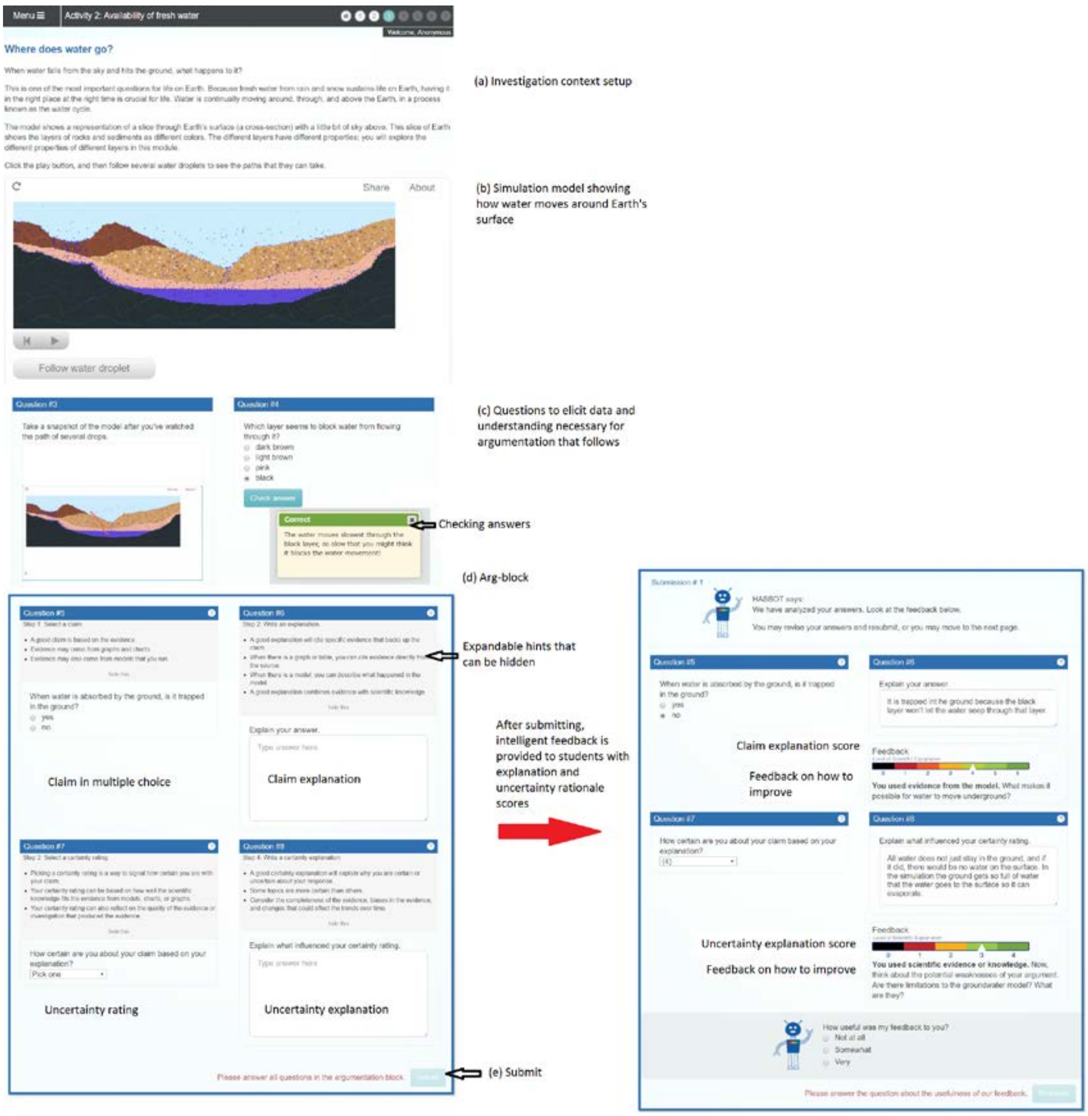

Figure 1. Groundwater Argumentation Task with Natural Language Processing-based Intelligent Feedback System.

Data collection and analysis

This study took place in two phases: identification of uncertainty attribution patterns based on students' written responses to argumentation prompts and uncertainty discourse impacted by intelligent feedback. In the first phase, we used written responses from 840 students to the groundwater argumentation task. These students were taught by 15 teachers in 8 states across the U.S. These teachers found the water module from various outreach efforts and voluntarily participated. They attended a summer workshop prior to implementing the module in their classrooms Students were in eighth to twelfth grades. Both genders were equally represented; 15% spoke English as second language; 80% used computers regularly for science learning prior to the module. We used the uncertainty attribution framework proposed by Kahneman and Tversky (1982) that listed four different types of uncertainty attribution in natural language: introspective confidence vs. personal theories under internal attribution and frequency-based distributional vs. singular causal under external attribution. Once we identified patterns in students' open-ended responses to explanation and uncertainty attribution prompts in the four-part scientific argumentation task, we conducted natural language processing (NLP) of these students' responses based on machine learning algorithms through c-rater-ML™ developed by Educational Testing Service. We also developed a feedback statement unique to each pattern of uncertainty attribution to address shortcomings.

A new version of the water module was developed by integrating automated scoring models into all eight arg-blocks so that students could receive feedback immediately after submitting their arguments. As soon as students clicked the submit button at the end of each arg-block, students' open-ended responses to the explanation and uncertainty attribution prompts were processed by the automated scoring engine developed to recognize whether they included a scientifically valid claim, evidence, and reasoning for an explanation (score ranging from 0 to 7) and how they articulated their uncertainty rationale (score ranging from 0 to 4). For the uncertainty attribution prompt analyzed in this study, the human-machine agreement was measured at 0.83. The whole process of submitting, autoscoring, finding feedback matching the score, and displaying the score and feedback to the student took two to five seconds in real time.

The second phase of this study was based on the intelligent feedback system-enabled water module. The module was implemented by four high school teachers in two suburban and two rural schools located in KY, MA, NH, and PA. Approximately 156 ninth to twelfth grade students used the water module. We recorded videos of the computer screens of 14 student groups, including student voices, as they worked through the water module. Uncertainty discourse was examined with these 14 student groups who worked on the groundwater model argumentation task shown in Figure 1. A total of 214 minutes of the videos were transcribed verbatim. We selected the groundwater argumentation task in this study because the task was the first time when students had to articulate uncertainty associated with model-based argumentation.

Findings

Uncertainty attribution patterns

From the uncertainty responses we analyzed, five different types of uncertainty attribution emerged. When asked to explain their uncertainty rating of their claim based on evidence they used, some students appeared to reiterate their uncertainty rating such as "We are not completely positive," "I'm kind of guessing but it makes sense and I'm pretty sure it's right," "I tried to think with common sense," and "I am almost certain about my answer." These statements describe introspectively the students' knowledge state about the claim they were making without any external reference to the phenomenon or personal justification rationale. On the other hand, we noticed a sizable number of responses referring to personal rationale based on their knowledge, abilities, or experiences relevant to understanding the driving question, processing and interpreting data, and conducting investigation. Examples of personal knowledge and ability references included "I didn't quite understand the rain example [shown in the model]," "I am familiar with the water cycle," "I didn't understand the question," "I cannot see the images very well," and "Based off my knowledge of what I learned about precipitation." This category also included personal experiences such as "Because my teacher told me," "We saw a movie about water being pumped," and "If the water were trapped, we wouldn't have enough water to live." We also categorized misconceptions about the groundwater topic ("Because most water travels through lakes and rivers") as personal rationale.

We found three general types of external uncertainty attribution. The first type simply acknowledges a scientific data source such as "Background information and the model," "The graph clearly, obviously, and very blankly shows this idea," and "After observing the diagram for a few moments, I managed to reach a conclusion as to what each dot represented and when it would change into the other dot." In these cases, students mentioned the source without providing details that explain the phenomenon or describe what happened in the data they were citing. The second type relates to external scientific disposition that either explains a mechanism for why water is not trapped ("We know that the water can pass every sediment layer but the black layer. The water moves though the layers easily on its way down to Earth until it hits the black layer which will stop the water from proceeding") or describes multiple outcomes related to water droplets ("The water moves though the layers easily on its way down to Earth until it hits the black layer which will stop the water from proceeding. Then after the water piles up and is overflowing to the top, evaporation occurs for some of them"). The former relates to singular mechanistic attribution and the latter relates to distributional attribution as multiple outcomes are acknowledged. We grouped both singular mechanistic and distributional attribution accounts under the external scientific disposition because these attribution types address the investigation at hand. The third type shows scientific limitations beyond the current investigation based on the groundwater model. Both ontic and epistemic uncertainty attributions were observed. The ontic attribution statements include "earth has many layers and not all of them can stop the water flow and some layers absorb the water and dispose the unwanted use of the water." The epistemic attribution statements focus mostly on model limitations such as missing factors ("because if they were they would stop flowing. However layers further down may be able to stop it from flowing").

Table 1 summarizes five categories we identified from students' responses to uncertainty attributions. In order to incorporate all that emerged from our analysis, we modified the framework of Kahneman and Tversky (1982) by adding a new category of external scientific source acknowledgement and external scientific limitation. Based on these characterizations, we created an ordinal progression by assigning scores from 0 to 4. The order of progression moves (1) from not mentioning to mentioning attribution, (2) from internal to external attribution, and (3) from vague external description to external scientific disposition then to scientific limitation. Based on this scoring method, we developed an intelligent feedback statement for each score.

Uncertainty discourse aided by intelligent feedback

In order to study uncertainty attribution discourse, we examined the discussions of 14 student groups, which were captured on video. Each group consisted of two or three students who responded to argumentation prompts in the arg-block together. These 14 groups made an average of 1.71 argument revisions after receiving real-time, intelligent feedback. Four groups did not make revisions; five groups revised once; three groups twice; two groups three or more times. Three groups claimed that groundwater was trapped while the other 11 groups claimed it was not. The claims did not change throughout revisions. When first submitted, explanations of nine groups included scientific reasoning that showed scientific mechanisms regarding whether the water was trapped or not. For example, one group wrote, "Water that is in the ground does not stay trapped in the ground because roots from plants suck up the water and through transpiration it evaporates from the plants to the atmosphere." Four groups included data they observed from the groundwater model: "It is trapped in the ground because the black layer won't let the water seep through that layer." Only one group's explanation did not explicitly mention data or reasoning: "The water stays always flowing." Intelligent feedback prompted those groups who included reasoning to also include data from the model and those who included data to also include reasoning. Six groups revised their explanations so that all 14 groups included at least data and/or reasoning at the end. Students' uncertainty rating ranged from 2 to 5 with an average of 3.8. Uncertainty ratings were rather stable before and after revisions as only one group changed their uncertainty rating to be higher (i.e. more certain) after revision. In initial submissions, six groups used internal attributions. Among the eight groups who used external attributions, four acknowledged a scientific data source without elaborating while four groups used external disposition based on singular mechanistic or distributional frequency-based accounts. Nine groups opted to revise uncertainty attribution: five groups' revisions resulted in external scientific disposition. Three groups reached the external scientific limitation level that discussed factors not currently represented in the groundwater model and that could alter their claims about whether the water was trapped in the ground.

From the video analysis, we identified several ways in which intelligent feedback supported students' recognition of uncertainty sources. First, feedback helped students frame uncertainty, which was understandably a novel task. Receiving low scores in uncertainty attribution was an eye opener to most students, which led to discussion and planning for what to do. For example, a group of students submitted uncertainty attribution by writing, "We are fairly certain of our answer because we watched many droplets come down and the path we chose was the fastest. We were also able to think out our answers reasonably." They received a score of 1, which meant personal attribution.

S1: A one? And a one?

S2: Huh? How'd we get a one for that?

S1: We were fairly certain...*[after reading the feedback]* what are you certain about from the groundwater model?

Table 1. Categories of uncertainty attribution

Source of Uncertainty	Categories	Description of categories	Intelligent feedback statement
No information (score 0)	No response	• Did not respond to the related uncertainty item but answered the linked claim and explanation items	You haven't explained your certainty rating. Have you compared the strengths and weaknesses of the evidence that you used to support your claim?
	off-task response	• "I do not know" or similar answers • Provided off-task answers	
Internal: Introspective confidence (score 0)	Restatement	• Restated the uncertainty rating	
Internal: Rationale (score 1)	Question	• Did/did not understand the question	Your personal beliefs, experiences, and attitudes can influence your certainty rating. How do the strengths and weaknesses of the scientific evidence affect your certainty rating?
	General knowledge/ ability	• Did/did not possess general knowledge or ability necessary in solving the question • Did/did not learn the topic (without mentioning the specific topic) • Can/cannot explain/estimate	
	Lack of specific knowledge/ ability	• Did not know specific scientific knowledge needed in the item set	
	Difficulty with data	• Did not make sense of data provided in the item	
	Authority	• Mentioned teacher, textbook, and other authoritative sources	
External: Scientific source acknowledgement (score 2)	Nominal data source	• Just acknowledged the existence of "data," "model," "chart," etc.	You mentioned that either the data or the model affected your certainty rating. Can you be more specific about how the data or model influenced your rating?
External: Scientific disposition (score 3)	**Singular mechanism**	• Referred to/elaborated knowledge or data that can explain the claim with data	You mentioned specific evidence and knowledge that influenced your certainty rating. Have you also considered the strengths and limitations of the data and models related to this question?
	Distributional frequency	• Compared and contrasted multiple cases	
External: Scientific limitation (score 4)	**Ontic uncertainty**	• Elaborated why the scientific phenomenon addressed in the item is uncertain	You recognized strengths and limitations of knowledge and evidence related to the current investigation. Excellent!
	Epistemic uncertainty	• Recognized the limitation of data provided in the item and suggested a need for additional data. • Mentioned that not all factors are considered • Mentioned that current scientific knowledge or data collection tools are limited to address the scientific phenomenon in the task	

S2: We were certain that the other one runs slower. We did well on the first one *[explanation score 4]*.

S1: I don't understand why we got through that *[explanation prompt]* really well.

S2: Then, we said, we said something else. We said that we…

S1: Hey, we're answering a completely different question than what it asked. That's why. [*Went back to the model to re-examine their evidence.]* It is asking us why we are certain about how the groundwater can get back up and be evaporated if it's not trapped!

S2: Yes!

S1: Are you certain of your answer?

S2: Oh, okay we figured it out.

S1: We are fairly certain, uh, wait, hold on….maybe it is trapped.

S2: No, it's not. We got a good score on explanation.

S1: Yeah, I know but it's not trapped…. This is what it is asking. It's asking, we answered, why we thought it is not trapped.

S2: Then, how do you explain it?

S1: Okay, we are certain or we are fairly certain because… um, 30% of the water we get is groundwater. And in the model it showed that water was being evaporated afterwards. Also, in the model, it showed the water evaporating from sediments.

[Resubmitted uncertainty attribution and received a higher score related to external singular mechanistic attribution.]

Second, students became more deeply engaged with interpreting data in light of their knowledge after receiving feedback. Students in the example above went back to the model to reexamine their claim; they also elicited a piece of knowledge that could be useful in interpreting data. The information that "30% of the water we get is groundwater" was learned earlier in the water module. Students voluntarily elicited this piece of knowledge to justify that water could not be trapped forever in the ground if they were to use groundwater in their life.

Third, feedback helped students revise their uncertainty attribution. We illustrate an example of a student group who first claimed that the water is trapped because "the water moved slowest through the black layer, so slow that you might think it blocks the water movement." With that explanation, the group chose an uncertainty rating of 4. The following sequence of uncertainty attribution occurred:

1. Initial attribution: "because we had an activity that backed up our reasoning."
2. Since this attribution was based on personal experience, feedback was given to students: "Your personal beliefs, experiences, and attitudes can influence your certainty rating. How do the strengths and weaknesses of the scientific evidence affect your certainty rating?"
3. First revision of uncertainty attribution: "because we had an activity that showed water sitting on top of the black layer which caused [us] to make the conclusion that the black layer absorbed water and it could not absorb any more so the water just sat on top."
4. The revision included their description of one outcome they observed in the groundwater model, the automated scoring engine recognized this statement as external, scientific disposition based on singular mechanism, thus the following feedback was provided: "You mentioned specific evidence and knowledge that influenced your certainty rating. Have you also considered the strengths and limitations of the data and models related to this question?"
5. In responding to this new feedback, the second revision was made: "because we had an activity that showed water sitting on top of the black layer which caused [us] to make the conclusion that the black layer absorbed water and it could not absorb any more so the water just sat on top. There are however limitations to the groundwater model because there are no plant life or roots in the ground that would help absorb some of the water." Note that students used a factor that was not represented in the current model to illustrate limitations in the model in making a claim about the groundwater trapping.
6. The second revision was recognized as external scientific limitation related to epistemic uncertainty. As such, congratulatory feedback was given to students: "You recognized strengths and limitations of knowledge and evidence related to the current investigation. Excellent!"

This revision sequence illustrates how students used intelligent feedback to turn their attention from internal to external sources of uncertainty and from focusing on finding an exemplar from the current data to limitations in the model where the data were generated.

Significance

While the pressure for implementing the Next Generation Science Standards is mounting, integration of science practices into current teaching and learning in science classrooms appears difficult. As students engage in science practices independent of one another, how to support students' diverse needs becomes an important issue in the design of instructional support systems. When students engage in argumentation with model-based evidence, uncertainty is prevalent as data and evidence are not fully understood by them or are not fully representing the phenomenon under investigation. The intelligent feedback system we tested to promote scientific argumentation delivers immediate, tailored supports for individual students commensurate with their progress on argumentation. Preliminary findings indicate that this automated feedback system can be seamlessly integrated into an online curriculum module to support students' uncertainty articulation about complex systems as part of written argumentation tasks.

References

Ben-Haim, Y. (2014). Order and indeterminism: An info-gap perspective. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 157–176). London: Routledge.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*, 26–55.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*, 303–352.

Boumans, M., & Hon, G. (2014). Introduction. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 1–12). London: Routledge.

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*, 473–493.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.

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.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*(2), 143–157.

Kandlikar, M., Risbey, J., & Dessai, S. (2005). Representing and communicating deep uncertainty in climate-change assessments. *Comptes Rendus - Geoscience*, *337*(4), 443–455.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810–824.

Lee, H.-S., Liu, O. L., Pallant, A., Roohr, K. C., Pryputniewicz, S., & Buck, Z. E. (2014). Assessment of uncertainty-infused scientific argumentation. *Journal of Research in Science Teaching*, *51*, 581-605.

Mcneill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*, 153–191.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, *94*(2), 203–229.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: National Academies Press.

Nussbaum, E. M., Sinatra, G. M., & Owens, M. C. (2012). The two faces of scientific argumentation: Applications to global climate change. In M. S. Khine (Ed.), *Perspectives on scientific argumentation: Theory, practice, and research* (pp. 17–38). New York: Springer.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*, 448–484.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, *12*(1), 5–51.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*(6048), 1393–1400.

Staley, K. W. (2014). Experimental knowledge in the face of theoretical error. In M. Boumans, G. Hon, & A. C. Petersen (Eds.), *Error and uncertainty in scientific practice: History and philosophy of technoscience* (pp. 39–56). London: Routledge.

Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. New York: Cambridge University Press.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under Grant Nos. DRL-1220756 and DRL-1418019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.