International Society of
the Learning Sciences

Students' Justifications for Epistemic Criteria for Good Scientific Models

Danielle Murphy, Ravit Golan Duncan, Clark A. Chinn
dm880@scarletmail.rutgers.edu, ravit.duncan@gse.rutgers.edu, clark.chinn@gse.rutgers.edu
Rutgers University, New Brunswick, NJ

Joshua A. Danish, Cindy E. Hmelo-Silver, Zachary Ryan, Morgan Vickery, Christina Stiso
jdanish@indiana.edu, chmelosi@gmail.com, zryan@iu.edu, moravick@iu.edu, cstiso@indiana.edu
Indiana University, IN

Abstract: Scientists use epistemic criteria to evaluate products of scientific inquiry, such as models. Engaging students with epistemic considerations as part of scientific practice is a growing focus of science education. Recent research has shown that students are able to identify and describe criteria for good scientific models. However, we know little about how students reason about and justify why specific criteria are important and which criteria are more important. In this study we explore these questions using data from interviews with 5th grade students. Interviews were conducted after a five-week model-based inquiry intervention and focused on students' reasoning about a set of criteria for good models developed by the class. Our findings illustrate the array of justifications provided regarding the importance of different criteria in their own right and relative to other criteria. We discuss implications for supporting students' use and evaluation of epistemic criteria in scientific practice.

Introduction

Recent reforms in the US have emphasized engagement with the practices of science and engineering as central to scientific literacy (NGSS Lead States, 2013). Competent engagement in practices involves both "doing" the practice and understanding what it means to do the practice (e.g., constructing models or explanations) (Barzilai & Chinn, 2018). Epistemic competence includes engaging in reliable cognitive processes to achieve epistemic aims that meet valued epistemic criteria (Barzilai & Chinn, 2018). These epistemic criteria are, "…the standards scientists use to evaluate the validity and accuracy of scientific products such as models, arguments, and evidence" (Pluta et al., 2011, p. 1). Examples of epistemic criteria for good scientific models or explanations include: models or explanations should fit the evidence, be conceptually coherent, fit with other established theories, and be parsimonious (Kuhn, 1977; Longino, 2002). These criteria become the guidelines for how to create, refine, and evaluate models. For example, models that fit with evidence have greater validity than models that are unsupported. Scientific communities have developed epistemic criteria that are shared and justified in terms of why they are important. Scientists may prioritize some criteria over others in pursuit of an aim. For example, a model's fit with evidence may be viewed as more critical than how parsimonious it is. Prioritization of epistemic criteria changes over time and may vary by discipline or even area of research. In general, however, criteria use is a core aspect of scientific practice across disciplines (Longino, 2002).

In this study we focus specifically on the practice of scientific modeling and the use of epistemic criteria to evaluate models. Giere (2004) defined models as, "…idealizations that scientists use to represent aspects of the world for specified purposes" (p. 742). Examples include the Copernican and Ptolemaic models of the solar system and Bohr's plum-pudding model of atoms (Frigg et al., 2020). Models are useful because they enable scientists to explain and predict natural phenomena. The process of building, testing, and refining models occurs within a community of practice in which scientists critique each other's ideas. Oreskes (2019) argued that through community argumentation, justification, and critique scientists can revise and build better epistemic products. As noted above, a key piece of the social construction of knowledge is deploying the criteria developed and accepted by the scientific community to critique and evaluate the epistemic products created by its members.

Given the centrality of epistemic criteria in scientific practice, we have argued that it is important to engage students in the development and use of epistemic criteria in science learning (Pluta et al., 2011). Developing, using, and justifying epistemic criteria provides two core benefits. First, it helps students develop a better understanding of scientific knowledge production and how scientists evaluate epistemic products (models, arguments, evidence) and decide between competing claims/models. Such understandings are crucial for a scientific literacy that espouses an informed trust in science. Second, similarly to how they function in the scientific community, epistemic criteria can be used by communities of learners to evaluate their own epistemic products. In this sense criteria can function as scaffolds to guide the production and evaluation of models. Given these arguments for the benefit of epistemic criteria use, the obvious question is whether students are able to productively engage with the development and use of such criteria.

Research suggests that students can do this. Pluta et al., (2011) found that seventh graders were able to generate a wide variety of epistemic criteria for good models and that the criteria align with those proposed by scientists. Av-Shalom et al., (2018) found that seventh graders grew in their ability to recognize and use fit with evidence as an epistemic criterion, which is a central criterion in modeling. However, research that investigates younger students' epistemic considerations when evaluating claims and evidence is sparse (Sandoval et al., 2014), and much of the current research has focused on students' description and use of shared criteria (e.g., Ryu & Sandoval, 2012). We know little about how students justify why these criteria matter, which criteria matter in different situations, and whether some matter more than others. When students talk about such justifications of criteria, they are engaged in *meta-epistemic discourse*, which involves explicit discussions about these criteria and the reasons why they matter (Chinn et al., 2020). Our study therefore focused on the kinds of justifications students provided for why specific criteria are important, and how multiple criteria relate to each other and are prioritized as a set (i.e., ordering within the set). Our research questions were as follows:

1. What criteria do fifth grade students develop for model quality?
2. How do students prioritize criteria?
3. What justifications do students give for why individual criteria are important?
4. What justifications do students provide for reasoning about criteria sets and order within a set?

Method

Context and participants

The study was conducted in a fifth-grade classroom during a five-week model-based inquiry unit with 50 students. During the unit, students reviewed evidence in the form of reports and simulations to develop a model that explained why fish were dying in a local pond (a eutrophication phenomenon). Students made models in a new modeling software called MEME (Model and Evidence Mapping Environment) that enabled them to create components and link them with arrows (mechanisms) to provide a causal account. Students could link evidence (the reports and simulations) to the various elements of the model.

The unit began with an activity in which students developed a shared class list of criteria for good models. To develop the class criteria list, students reviewed a series of model pairs, determining which model in each pair they believed to be the better model and why. Based on these evaluations, students individually made a list of the most important criteria for good models and ranked the order of importance. Through a class discussion the teachers drew on students' suggestions for criteria to develop a class list of five shared and agreed-upon criteria for good models. The class criteria list was revised once throughout the unit. Thus, students had opportunities to use the criteria as they developed and revised their models as well as when they critiqued peer models.

The unit was taught in a public charter school in the Northeast of the United States. The school performs above average on standardized state tests on Mathematics, Language Arts and Science. The demographics of the school were 60% White, 22% Asian, 9% African American, 7% Latinx, and 2% other; 5% of students were eligible for free and reduced lunch.

Data collection and analysis

The data for this study are from interviews that were conucted at the end of the intervention. Sixteen students were interviewed in dyads. The interview consisted of a series of questions to assess students' conceptions of criteria for good models, including criteria importance, order, and justifications. In the interview, students were briefly shown their class's final criteria list and then were given the individual criteria and asked to order the criteria from most to least important. As students were ordering the criteria list, they had the option to add any additional criteria they felt were not covered by the class list. Students were then asked to provide justifications by explaining why they chose the order that they did, beginning from the most important and working their way down to the least important. They were also asked if any criterion was way less or way more important than the others or if they were all similar in importance.

The interviews were videotaped, audio recorded, and later transcribed. The transcripts were then reviewed to identify instances in which students provided reasons about why criteria are important, or why one criterion was more important than another. Through constant comparison (Glaser, 1965), these statements were categorized based on the kinds of justifications they presented. We discuss these categories in the next section.

Results

Research Question 1: What criteria do fifth grade students develop for model quality?

As noted above, students generated a criteria list early in the intervention and then revised it later. The final list is shown in Figure 1.

1. Be organized, clear, and explain what is happening
2. Include relevant evidence
3. Include evidence to support both components and mechanisms
4. Include thorough explanations of the evidence (should completely explain what the evidence shows or why it is relevant
5. Not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)

Fig 1. Class Criteria List

It is important to note that while there are five criteria in the final list, some of them include multiple ideas that arguably could have been separated into separate criterion. For example, the criteria "*be organized, clear, and explain what is happening"* includes three separate ideas or propositions: be organized, be clear, and explain what is happening. This proved to be important for our analyses because during the interviews students sometimes referred to one, two, or all of the propositions within a criterion, and dyads differed in how they treated criteria in this sense. For example, some dyads referred to particular criteria as a whole (e.g., "*be organized, clear and explain what is happening"* treated as a single criterion), whereas other dyads referred to only one proposition within the criterion (e.g., "*be organized* or *explain what is happening"*). There was also redundancy, as some ideas appeared in multiple criteria. For example, the notion of evidence appeared, in some form, in four out of the five criteria: evidence needs to be relevant (2), it needs to support both components and mechanisms (3), it should be thoroughly explained (4), and there should not be extra evidence (5).

In a prior study of individual seventh graders' development of lists of criteria, Pluta et al., (2011) classified students' criteria into five broad categories: (a) criteria that articulated the goals of models, such as explanation; (b) criteria related to evidence fit, such as quality and quantity of support; (c) criteria that specified constituent parts that needed to be present, such as diagrams; (d) criteria related to communicating the model clearly, such as being clear and organized; and (e) criteria related to other epistemic features such as accuracy. The criteria in the 5th grade class list in this study fall into four of the five categories: (a) goals of models (explain what is happening); (b) evidential criteria (include evidence to support both components and mechanisms); (c) criteria that specified constituent parts that needed to be present (not have extra or irrelevant components, mechanisms, or evidence); and (d) communicative elements (be organized and clear). There were no criteria related to the fifth category, epistemic features, from Pluta et al., (2011).

Research Question 2: How do students prioritize criteria?

In the interviews, students were asked to prioritize criteria from most to least important. Table 1 shows the class criteria list and how many dyads placed each criterion in the first to fifth location on the list. In our analysis, the placement of criteria was given a point value: A criterion in first place was awarded five points, a criterion in second place was awarded four points, and so on. These points were added to yield a "total points" score (shown in the final column) to give a sense of overall criteria prioritization among the dyads.

Table 1: Ranking of criteria and corresponding point value

Criteria	**1st**	**2nd**	**3rd**	**4th**	**5th**	**Total Points**
Include evidence to support both components and mechanisms	3	2	2	1	0	**31**
Be organized, clear, and explain what is happening	3	1	2	2	0	**29**
Include thorough explanations of the evidence (should completely explain what the evidence shows or why it is relevant)	1	2	2	1	2	**23**
Include relevant evidence	0	3	2	0	3	**21**
Not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)	1	0	0	3	4	**15**

Criteria prioritization reveals a few general trends. The criteria "*be organized, clear and explain what is happening"* and "*include evidence to support both components and mechanisms"* were ranked the highest. The criterion "*not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)"* was ranked lowest. The criterion "*include thorough explanations of the evidence (should completely explain what the evidence shows or why it is relevant)"* and "*include relevant evidence"* varied in placement across dyads with the largest spread throughout. However, these trends were weak as there was variation in placement across dyads. More information is given in the next section about why this variation may have occurred, as we analyze the justifications students gave for why each criterion was important.

Research Question 3: What justifications do students give for why individual criteria are important?

Students gave a variety of reasons for why each criterion was important. Table 2 shows the class criteria and the justifications students gave for why each criterion is important. The frequency counts are given after each justification type. For example, three dyads said that the first criterion ("*include evidence to support both components and mechanisms"*) is important because it makes the model believable. The frequency counts vary because in some cases students gave only one reason in support of a criterion, whereas in other cases they gave several reasons (this varied across criteria and dyads). Also, sometimes one student spoke for the dyad, whereas at other times both students spoke and either gave different justifications or elaborated upon each other's ideas. Elaborations of the same justification by both students within a dyad were counted as one instance. Elaborations that focused on a different aspect of a criterion or a different justification were counted as two separate responses. This resulted in an unequal number of justifications per criterion.

Table 2: Justifications for why individual criteria are important

Criteria	**Justification Types (Frequency counts)**
Include evidence to support both components and mechanisms	• Evidence makes the model believable, true or correct (3) • Evidence shows why something is happening, or it shows the thinking behind the model (2) • Evidence should support all components and mechanisms; all parts of the model (2)
Be organized, clear, and explain what is happening	• When a model is clear and organized it is understandable or readable (5) • The model will show what it's supposed to show (1) • People will be able to think how you think (1)
Include relevant evidence	• Relevant evidence means the parts of the model, idea and purpose are understandable (2) • Relevant evidence explains the model (1) • Without relevant evidence, no one will believe you (1)
Include thorough explanations of the evidence (should completely explain what the evidence shows or why it is relevant)	• Explaining the evidence is needed so people know what the model is showing, the main idea, why something happens (4) • Explanations show how the evidence supports the model (3) • Without explanations and evidence there is no model (1) • Evidence makes the model true- evidence proves the point (1)
Not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)	• Have to have real information (1) • Focus on what we need first; not having everything you need is worse than having everything you need plus extras (1) • Extra/irrelevant pieces aren't good- makes the model unfocused and off topic (1)

The reasons students gave for the importance of each criteria were varied. Due to space constraints, the justifications for two criteria will be more fully discussed here. The first criterion ("*include evidence to support both components and mechanisms"*) was selected because it was the highest ranked and because evidence fit is a centrally important epistemic consideration. Students valued evidence because they viewed it as making the model accurate. Students referred to accuracy through the terms, "believable," "true," or "correct." Students also valued evidence because they viewed it as showing why something is happening. This implies that students understood the role of evidence in making the model valid, accurate, and complete. Finally, students specified that evidence must support all parts of the model, not only one piece. This shows that students understood that models are not a unitary entity but have constituents that all need to be supported individually. Their justifications for this criterion show that students valued fit with evidence in general because it confers accuracy and validity, as well as in more nuanced ways—in terms of the need to specifically support each and every part of the model.

We discuss the second criterion ("*be organized, clear, and explain what is happening"*) because out of all the justification types, the idea of understandability was stated most frequently and was applied to justify this criterion. Five dyads explained that this criterion was important because a model that was clear and organized would be understandable. Students also said that this criterion was important because if a model is clear and organized, then it will achieve its purpose (show what it is supposed to show), and the audience will be able to follow your reasoning (think how you think). This shows an awareness of the communicative role of model representations and the importance of clearly representing one's ideas. These justifications also point to students' understanding that models have an epistemic purpose or aim that they need to achieve.

We wish to note that students cited understandability as a justification for four out of five criteria, which shows it was not only a popular justification for the criterion "*be organized, clear and explain what is happening"*; it was a popular justification for criteria in general. Students said that models that are organized and clear, include relevant evidence, show how evidence supports the model, and have fewer (and only needed) parts, will be understandable. Although the justification of understandability was given for different reasons, the frequency of this justification type shows that students privileged understandability as an important justification for criteria for good models but also used this justification in different ways (i.e., had different notions of what understanbilty means in relation to models).

Another interesting pattern was the variation in justifications students gave for why evidence is important. (These justifications spanned the four criteria that mentioned evidence.) Out of the 29 total justifications given, 20 pertained to evidence. Some dyads argued that evidence is important because it makes the model accurate (i.e., evidence makes the model true, correct, or believable). This supports the previous point that students are relating evidentiary support to model accuracy. Some dyads related evidence to understandability claiming that relevant evidence means that parts of the model (i.e., components and mechanisms) are understandable, whereas other dyads reasoned that relevant evidence makes the idea or purpose of the model understandable. Finally, some dyads thought that without evidence and explanations, there would be no model. These results show that students understood the importance of evidence in multiple distinct ways.

Research Question 4: What justifications do students provide for reasoning about criteria sets and order within a set?

Students also reasoned about criteria as a set and the order within the set. Students discussed criteria as a set by giving reasons for why certain criteria were more important than others by discussing criteria in relation to each other. This section extends the prior results by focusing on connections across and between criteria.

Table 2: Justifications about criteria sets and order within a set

Justification	Example
Encompass other criteria. A criterion is more important because it encompasses others.	This is first because it includes all the other criteria mixed into one.
Specificity. A criterion is more important because it is more specific.	These two are about the same but this one [irrelevant pieces] is more specific...
Essential to the model. A criterion is more important because it is essential to even having a model.	Without explanations and evidence there is no model.
Redundancy. A criterion is less important than others because it is redundant.	One is opposite of the other so they're kind of saying the same thing.
Already known. A criterion is less important because it is already known.	We put this one last because this is something you already know.

The first justification, encompass other criteria, was used to explain why one criterion was more important because it encompassed others. This is shown in the following excerpt from a transcript of Ivy and Arya explaining why they chose their first criterion "*be organized, clear, and explain what's happening*". Note that throughout this section, the descriptions in brackets clarify what the speaker was referring to:

Ivy: Umm, so we chose the order because like, because like you first want to make sure it's organized and clear 'cause like... if you... if that was like the last thing you would do, then um... you would like... then it wouldn't be that clear because then maybe like organizing clearly first is like the uh the hard thing to do, so you want to get it done first. Yeah.

Arya: To add on to what she said, if you have a model that's really messy, if somebody's trying to get a conclusion out of it, they're not going to be able to understand what's happening because they're going to get confused, and it's like more important for people to understand your model and for you to explain it thoroughly as if they don't know like what's happening.

Ivy: Yeah, and like these [all criteria apart from the first one] would just be extra checking, organized and clear.

Ivy and Arya explained that the first criterion ("*be organized, clear, and explain what's happening"*) is the most important because the model needs to be understandable. If the model is messy it will be confusing, and the reader will not understand what is happening. Therefore, the dyad concluded, all other criteria are basically encompassed by (or subordinate to) this first criterion.

Other dyads referred to the idea of some criteria being encompassed by others, but in a slightly different way, by claiming that one criterion was entailed by another. For example, Geona and Deborah argued that the fifth criterion ("*not have extra or irrelevant components, mechanisms, or evidence [all should help answer the question]"*) implies that relevant components and evidence are included and therefore entails the third criterion ("*include relevant evidence"*).

Deborah: It ["*not have extra or irrelevant components, mechanisms, or evidence (all help answer the question*)"] included already like relevant evidence.

Geona: Yeah like "*extra and irrelevant components*" it shouldn't have that so that means it should include relevant evidence.

The second justification privileged specificity; that is, more specific criteria are better. This rationale was given by John and Miguel for why the criterion "*not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)"* was ranked higher than the criterion "*include relevant evidence."*

John: "*Not have extra or irrelevant components, mechanisms, or evidence."* I still think it's the same as this "*include relevant evidence"*. It's just saying don't have extra irrelevant evidence so it's basically the same. Don't have uh, have relevant evidence and then mechanism and components, that's a little more detailed than this one so we put this one ahead.

John and Miguel reasoned that these two criteria are basically the same, but the prior criterion was more specific and detailed and was therefore ranked higher.

The third justification type was that some criteria are important because what they specify is essential to models and modeling. This was shown in the justifications given by two dyads. The first dyad, Geona and Deborah, stated that the criterion "*include thorough explanations of the evidence (should completely explain what the evidence shows or why it is relevant)"* was the most important because if a model doesn't have explanations and evidence, then there is no model:

Deborah: Umm because here the most important thing for a model is that it shows ummm it needs to show explanations cause without like explanations and evidence there's basically no model.

The students perceived explanations and evidence as being so integral to the model that it is not possible for a model to exist without them.

Another dyad, Patrick and Grant, explained that the criterion "*be organized, clear, and explain what is happening"* is the most important because if it is not organized the audience will not be able to read and understand it, rendering it pointless:

Patrick: Because organize, clear, and explain what is happening is the most important because if it's not organized then other people won't be able to read it so basically there's no point in doing it if it's not organized or clear.

This implies that clearly communicating ideas through the model is of paramount importance. If the audience are not able to read the model, the ideas within it will not be communicated or understood, therefore the purpose for making the model has not been attained.

The fourth justification type ranked some criteria lower because they were deemed redundant. By redundant, we mean that other criteria were more important and covered the most important standards for good models. Therefore, the criteria that are classified as redundant are a part of the set, but they are not as significant as others. For example, Laila and Olympia explained that the criterion "*include relevant evidence"* is the least important because a lot of the other criteria already show that evidence is important:

Laila:	Umm... we put this one last because we already have a lot of these that show that evidence is really important to have in a model and that it needs to answer the question so that it makes sense. So, including relevant evidence is sort of just like… including relevant evidence means that you need to include good and pretty good evidence, but I think these are more important because you need to have mechanisms and components and umm... you need to make sure your evidence connects with the question and you have to be organized and clear and explain what's happening.

Laila explained why the criterion "*include relevant evidence"* is the least important and showed why she placed other criteria higher. In her explanation, she explicitly listed all other criteria and says they were more important because models need to have components and mechanisms, make sure evidence connects to the question, be organized and clear, and explain what is happening. Laila felt that this criterion is least important because it has already been said in these other, more important criteria.

The fifth and last justification category ranked some criteria as specifying aspects of models students already knew and therefore the criterion was deemed less important. For example, the dyad James and Ethan said that the criterion "*not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)"* was least important because they already knew not to do this:

James:	And this one is pretty much commonsense.
Ethan:	Yeah.
James:	If something doesn't go with your model, then-
Ethan:	Don't put it there.
James:	Yeah, don't put it there.

James and Ethan reasoned that the criterion "*not have extra or irrelevant components, mechanisms, or evidence (all should help answer the question)"* is "commonsense" and "something you already know." As a result, the criterion was given less priority and was placed at the bottom of the list.

Discussion

Epistemic criteria are central elements of scientific practices, including the practice of modeling. Scientists evaluate models in accordance with epistemic criteria, such as fit with evidence and explanatory power. If we are to engage students similarly with epistemic criteria that they develop and are accountable to in the context of a learning community, we need to better understand how students understand these criteria, their importance, and their relative priority in relation to other criteria. Towards that end, we analyzed fifth grade students' justifications of why specific criteria are important and their reasons for rank ordering them. Our analyses show that students were able to develop a class criteria list to guide model creation and revision. Students were also able to evaluate criteria by prioritizing and reasonably justifying why criteria were important. Further, students were able to provide reasonable justifications about criteria as a set and criteria in relation to each other. Some of these justifications included important aspects of criteria lists, such as specificity, redundancy, essential elements, and being encompassed and entailed by others.

Our findings extend previous research, which showed that seventh graders could develop appropriate epistemic criteria for model goodness, by demonstrating that younger students share this competence, and further, that they can prioritize criteria. More significantly, our findings demonstrate that fifth graders have meta-epistemic competence. By this we mean that they are able to reason about and justify epistemic criteria for scientific models. First, students are able to describe the criteria of good scientific models. This is valuable because it means that students are aware of their thinking and are able to articulate ideas about epistemic criteria. This may seem trivial, but such discussions are rare in classrooms, and it is not simple to articulate meta-reasons for complex epistemic considerations (Chinn et al., 2020). Second, students are able to engage in a more sophisticated level of meta-epistemic discourse by evaluating criteria of good scientific models. This was shown by the students' justifications, which reveal that students understand why criteria are valuable. This is a novel finding.

Meta-epistemic competence is an important aspect of scientific practice, and we have demonstrated that elementary school students, as young as fifth graders, exhibit this competence. This is an impressive capacity that instruction can build on. Not only can students describe very reasonable criteria; they can also articulate why their criteria are important in ways that are sensible and compelling. Instruction can attend to these student competencies by engaging students in developing and applying criteria lists to models, and engaging students in meta-epistemic discussions of why these criteria matter, to whom, and under what conditions. In some cases, people differ in the criteria they value and this can lead to deep epistemic disagreement (Chinn et al., 2020). Resolving such disagreement entails reflecting on the criteria the different parties espouse and why. Discussions of this sort, about the relative merits of different criteria, are important yet rare in schools. If we want students to build competence in noticing differences in criteria use, and in arguing about their utility and value, we need to engage them in these kinds of discussions early and often. Our study has shown that even fifth grade students are capable, at least to some extent, of justifying criteria and reasoning about their relative merits.

References

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, *82*(4), 417-436.

Av-Shalom, N. A. Y., El-Moslimany, H., Duncan, R. G., & Chinn, C. A., (2018). Changes in students' use of epistemic criteria in model evaluation. In J. Kay & R. Luckin (Eds.), *Rethinking learning in the digital age: Making the learning sciences count, 13th International Conference of the Learning Sciences (ICLS) 2018 Volume 2* (pp. 768-775). International Society of the Learning Sciences.

Barzilai, S., & Zohar, A. (2016). Epistemic (meta) cognition: Ways of thinking about knowledge and knowing. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 409-424). Routledge.

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2020). Disagreeing about how to know: The instructional value of explorations into knowing. *Educational Psychologist*, *55*(3), 167-180.

Frigg, R., & Hartmann, S. (2020). *Models in science. The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/models-science/

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*, 742–752.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436-445.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. University of Chicago Press.

Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, *84*(1), 71-94.

Longino, H. E. (2002). *The fate of knowledge*. Princeton University Press.

Oreskes, N. (2019). *Why trust science?*. Princeton University Press.

Pluta, W. J., Chinn, C. A., Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching, 48(5), 486-511.*

NGSS Lead States. (2013). *Read the Standards. Next Generation Science Standards: For States, By States.* https://www.nextgenscience.org/search-standards

Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, *96*(3), 488-526.

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014) Developing children's early competencies to engage with science, *Educational Psychologist, 49* (2), 139-152.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1761019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

