Students' Use of Evidence and Epistemic Criteria in Model Generation and Model Evaluation

Ravit Golan Duncan, Carol Tate, and Clark A. Chinn
Rutgers University,10 Seminary Place, New Brunswick, NJ 08901
ravit.duncan@gse.rutgers.edu, carol.tate@gse.rutgers.edu, clark.chinn@gse.rutgers.edu

Abstract: The Next Generation Science Standards and the Framework for Science Education emphasize the importance of engaging learners with the core scientific inquiry practices of modeling and argumentation. Students are also expected to understand the epistemic grounds and norms that accompany these practices. We report on a study in which we engaged middle school teachers, and their students, in model-based inquiry, with particular emphasis on developing models and evaluating competing models using evidence. Analysis of students' written arguments, in the context of an assessment task in genetics, suggests that students use both secondary epistemic criteria, relating to communicative features of models (labels, drawings), as well as primary epistemic criteria, relating to evidence-model fit. Most students used at least one, and often several, provided pieces of evidence to support their arguments. We also discuss some instructional implications and tradeoffs in selecting evidence for such model generation and model evaluation tasks.

Introduction

The recently released Next Generation Science Standards (NGSS) (Achieve, 2013) adopt a view of scientific inquiry as a knowledge-building enterprise that employs a systematic, and evidence-based, approach to building models that explain the world around us (Giere, 2004; Godfrey-Smith, 2006). Scientific models are abstract, simplified representations of important aspects of the scientific phenomenon under study (Doerr & Lesh, 2003). These models are developed iteratively through a process of testing and revision. Evidence, and reasoning about evidence, is at the core of these processes (Longino, 2002). Scientists developing these models operate within a community of research, with continually negotiated norms regarding what counts as good evidence, arguments, and models (Kitchner, 1993; Kuhn, 1977; Latour, 1987).

Over the past two decades, there has been substantial research investigating the learning and teaching of scientific modeling practices (e.g. Driver, Leach, Millar, & Scott, 1996; Grosslight, Unger, Jay, & Smith, 1991; Lehrer & Schauble, 2000; Schwarz & White, 2005; Treagust, Chittleborough, & Mamiala, 2002). However, we still know relatively little about the ways in which students understand what counts as a good model, evaluate the quality of evidence, and relate evidence to one or more explanatory models (Lehrer & Schauble, 2000; Pluta, Chinn & Duncan, 2011; Schwarz *et al*, 2009). For example, Schwarz and White (2005) had students evaluate models using four criteria that were provided to them: accuracy, plausible mechanism, consistency, and utility of models. They found that students who used the four criteria showed a better understanding of the nature of modeling, scientific inquiry, and the targeted physics content compared to students who completed the same instructional unit, but without explicit use of criteria. In our own work we have shown that middle school students are also capable of generating their own criteria for judging model quality (Pluta, et al., 2011). Students' criteria were predominantly about communicative features of models such as models being clear, labeled, organized, and including pictures. However, almost a quarter of the students did note criteria that related to the model's fit with evidence, and almost half claimed that models should explain the phenomenon under study. These findings are congruent with earlier research suggesting that, at least some, high school students see models as important for developing explanations and making predictions, and that they are revised in light of evidence (Grosslight et al., 1991; Treagus et al., 2002). Helping students develop more sophisticated epistemic understandings of models and their role in science is important if we want them to fully grasp and engage with the modeling practices advocated by the NGSS and the Framework for Science Education (National Research Council, 2011). In accordance with the ICLS 2014 theme, developing such epistemologies is part and parcel of learning science and becoming scientifically literate.

In our current project, we have developed a set of epistemic scaffolds to support student engagement with models, evidence and the relationship between them (Chinn & Buckland, 2012; Rhinehart, Chinn & Duncan, in press). There are three core scaffolds that we have used with middle school teachers and students: (a) student generated lists of criteria for model-goodness; (b) Model-Evidence-Link diagrams (MELs), in which students use five different types of arrows (support, contradict, strongly support, strongly contradict, and irrelevant) to connect each piece of evidence to multiple competing models; and (c) evidence rating boxes within the MEL diagrams, in which students record their judgments of the quality of each piece of evidence on a scale from 0 (very poor evidence) to 2 (high quality evidence).

Here we report findings from a modeling-and-argumentation assessment task in which students had to

first develop their own models of a hypothetical genetic disorder given a few pieces of evidence. They then used a MEL diagram to evaluate, and choose between, two competing explanatory models of that disorder in light of additional provided evidence. Students completed this assessment at the end of a 5-week unit on genetics. Our research questions thus align with our two-fold goal for the task: (a) evaluating the extent to which students can generate mechanistic models of the cellular and molecular mechanisms that underlie genetic phenomena, and (b) evaluating students arguments and, in particular, their use of evidence and epistemic criteria.

Research suggests that students struggle to provide mechanistic accounts of genetic phenomena that explain how our genetic information brings about physical traits (Lewis & Kattmann, 2004; Marbach-Ad & Stavy, 2000). This is, in part, due to the many unfamiliar cellular and molecular entities involved, such as DNA and proteins (e.g. Lewis & Wood-Robinson 2000; Marbach-Ad, 2001;Venville & Treagust, 1998), and in part to the current instructional methods that tend to blackbox the protein-based mechanisms that link genes to traits (Duncan & Reiser, 2007). Several researchers have proposed instructional frameworks and scaffolds to support students in developing mechanistic explanations of genetic phenomena (Duncan & Reiser, 2007; van Mil, Boerwinkel, & Waarlo, 2011). Implementations of such scaffolded curricula have met with some success at the middle and high school level (Duncan & Tseng, 2011; Duncan, Freidenreich, Chinn & Bausch, 2011). Introducing these ideas early on is the key to supporting more robust understandings by the end of schooling. This view is reflected in the NGSS, which, unlike prior iterations of the standards (NRC, 1996), introduce the relationship between genes, proteins, and traits at the middle school level (MS-LS3-1).

Therefore, as part of our study of modeling and argumentation with middle school students we have developed a five-week unit in genetics that focuses predominantly on Mendelian genetics but also addresses the link between genes, proteins and traits in the context of genetic resistance to HIV (described in detail below).

Methods

Study Context

The study was conducted in a relatively large suburban 6th and 7th grade middle school (approximately 1450 students) in the Northeast. The majority of the students in the school were Caucasian (61%) with a large minority of Asian students (28%), and small minorities of Hispanic (6%) and African-American (5%) students. Approximately 14% of the students were eligible for free and reduced lunch. Four 7th grade teachers participated in the study with their approximately 400 students. Participating teachers implemented five months of instruction using materials we developed jointly, interspersed with their own materials. The study involved two conditions: (a) the treatment condition included a consistent and explicit focus on developing and using criteria for model-goodness that were student-generated and revised periodically throughout the duration of the implementation, and (b) the control condition in which there was no explicit and public focus on model-goodness criteria. Both conditions used the MEL diagrams and the evidence rating boxes.

The implementation study began with a set of activities designed to: (a) introduce students to the norms of argumentation discourse (giving reasons, disagreeing nicely, etc.); (b) engage students in the generation of a consensus list of model goodness criteria (*only* in the treatment condition); and (c) introduce students to the MEL diagrams and evidence rating boxes and procedures. These introductory activities were followed by a unit on cell organelles in which the teachers used study materials for two of the organelles they taught- chloroplast and nucleus (they used their own materials for other organelles typically covered in this unit). Following the cell unit was the 5-week genetics unit. This unit began with several lessons about Mendelian genetics during which students developed model for the "rules" governing inheritance patterns they observed in pedigrees, and then learned the relevant terminology and algorithms (Punnett squares) used to describe inheritance patterns and the probabilities of particular gene and trait combinations. The unit then turned to molecular genetics and the remaining lessons dealt with inherited resistance to HIV (some people are not susceptible to HIV infection because the virus cannot enter and infect their white blood cells). This part of the unit also included a set of teacher-planned activities about the structure and function of DNA. The HIV lessons involved the evaluation of two competing models for the genetic basis of HIV resistance that linked a mutation in a gene to the resistance trait. In one model the mutated gene gives instructions for making a novel protein that attacks the virus and thus confers resistance; in the second model the mutated gene results in a missing membrane protein that is normally used by the HIV as an anchor (necessary for infection of the cell by the virus). The second model approximates the currently acceptable mechanism for HIV resistance. As in most of our instructional activities, students evaluated these models against multiple pieces of evidence, such as a simplified summary of a scientific study about the presence or absence of particular proteins in the cell membrane of normal and resistant individuals. Students wrote extensive arguments in support of their chosen model; students in the treatment condition were encouraged to refer to the model-goodness criteria as they developed their arguments.

In this paper we report about one of the teachers and her 90 students: 40 in two class sections assigned to the treatment condition, and 50 in three class sections in the control condition. The teacher was untenured and in her second year of teaching. She held progressive views of teaching and was eager to engage her students

with model-based inquiry instruction. She enacted the genetics unit with high fidelity based on our field notes and video tapings of her lessons.

Data Sources and Analysis

The written assessment described herein was given to all students at the end of the genetics unit. The assessment was comprised of two tasks both involving a scenario of a hypothetical skin disease "DEB" in which individuals have blisters in their skin. In the first task, students were asked to explain using pictures and words, "…what you think is happening inside the bodies of people with DEB?" They were also provided with three pieces of evidence related to DEB: Evidence 1 described the inheritance pattern of DEB, Evidence 2 compared samples of healthy skin with skin from DEB patients, and Evidence 3 provided a diagram of healthy skin showing the layered structure. We analyzed students' models to ascertain whether they included an explanation that linked genes to the trait (blisters) via a protein-based mechanism, and whether they used the provided evidence in their models. Students were then asked to critique their model: "How good do you think your explanation is? Give at least four reasons for your answer." Responses to the self-assessment prompt were examined in terms of how students rated the quality of their models, the kinds of model-goodness criteria they referred to in their evaluation of the model, and whether they identified any shortcomings of their models.

In the second task, students were presented with two explanatory models of DEB that provided a mechanism linking a gene to the blistering (see Figure 1).

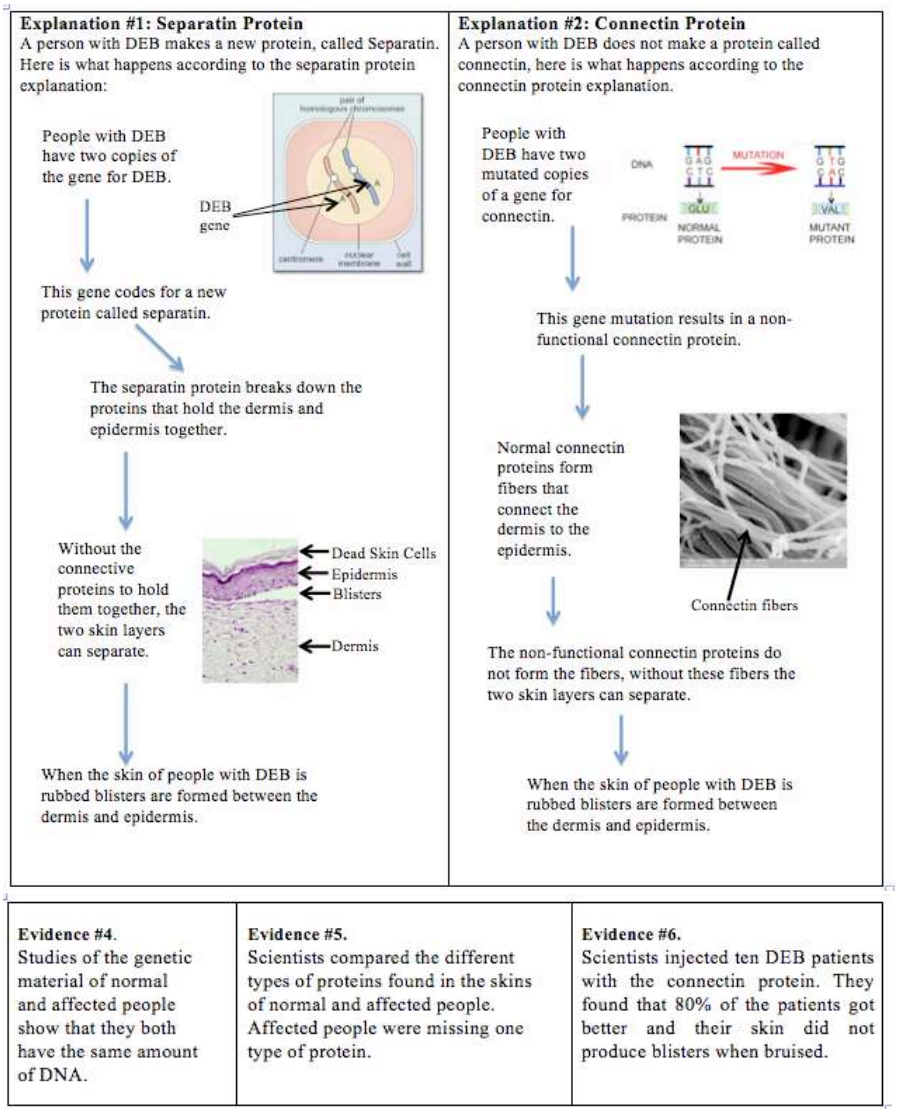

Figure 1. Model-Evaluation Task

The first model (the separatin model) postulated a mutated gene coding for a novel "separatin" protein that caused the skin layers to separate resulting in blisters. In the second model (connectin model) the mutated

gene results in the lack of a "connectin" protein that normally holds the skin layers together. The second model, lack of a protein, is the correct explanation of the real disorder on which the DEB scenario is based. Students initially chose which model they believed to be correct, and then read three additional pieces of evidence shown in Figure 1: Evidence 4 stated that both normal and affected individuals have the same amount of DNA (this evidence supports both models and is essentially irrelevant to choosing between them); Evidence 5 indicated that scientists discovered that affected individuals are missing one type of protein (this evidence was intended to support the connectin model); and Evidence 6 described a study in which scientists injected 10 affected individuals with connectin and 80% of the patients got better (this evidence supports the connectin model but has a rather small sample size). Students were then prompted to reconsider their choice of models in light of the evidence and write a reasoned argument to support their choice.

We analyze these arguments using coding schemes adapted from prior research (Dianovsky, Duncan & Chinn, 2013) to capture the quality of student arguments in terms of students' use of evidence including: (a) how did they interpret the evidence, (b) how many pieces did they cite, (c) did they explain the link between the evidence and the model, (d) did they address counterevidence (if they chose the incorrect separatin model), and (e) did they include any counterarguments against the competing model.

Results and Discussion

Students' Models of DEB

The student-generated models of DEB were mostly phenomenological and did not include any protein-based explanations of DEB (see Figure 2a). Only one of the 90 participating students provided a model that included a postulated mutation in a gene resulting in missing protein that would normally connect the dermis and epidermis skin layers preventing blistering (see Figure 2b). These results are fairly disappointing and we were surprised that none of the other students provided a mechanistic explanation. It may be that they did not generalize the role of genetic mutation and proteins in genetic phenomena from the HIV example taught in the unit. Our prior research did suggest the need for multiple examples and support in generalizing the gene-protein-trait schema (Duncan, et al., 2011), however, due to time constraints with the genetic unit in this study, we did not develop these additional activities. It seems that despite an emphasis on genes being instructions for proteins in the nucleus lesson (taught before the genetics unit), and a similar emphasis in the teacher-generated activities about DNA (central dogma), students did not develop a generalized schema that they could apply in other contexts. It is also the case that the evidence we presented in this first task (Evidence 1-3) did not deal with proteins and thus students were not compelled by the evidence to introduce proteins into their models.

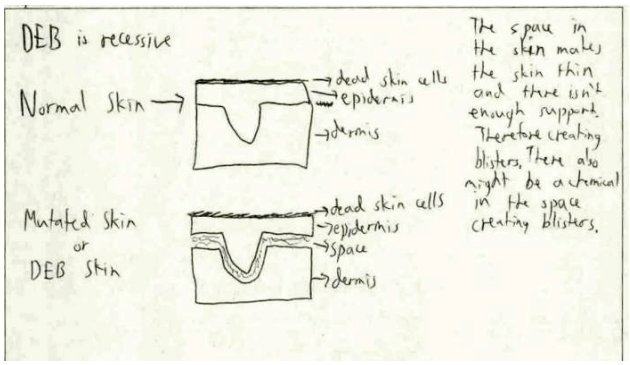

Figure 2a. Typical Student-Generated Model of DEB

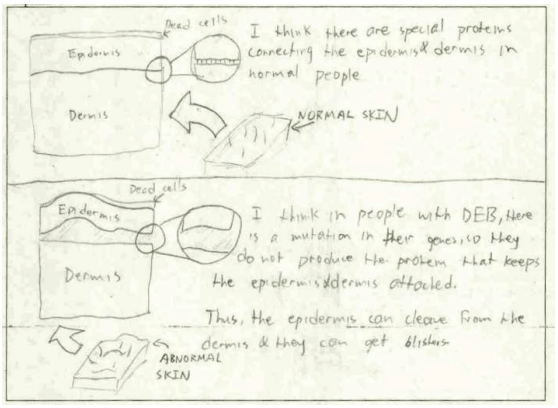

Figure 2b. Mechanistic Model of DEB

Given that students' models were not truly explanatory, we were interested in seeing how they evaluated their own models. We found that most students rated their models positively (noting that the model is very good, good, or OK). These students often cited several of their class model-goodness criteria in justification of their response, including clarity, having drawings, labels, and fit-with-evidence. Table 1 illustrates the criteria and frequency of citation by students. Note that the "I used the evidence" criterion was fairly prominent, and given that the evidence was at the phenomenon level, these students are sensibly citing the criterion. Many students essentially re-drew the evidence showing the two skin layers and the blisters. Thus for all intents and purposes they did use the evidence. Interestingly, there seem to be no significant differences between the treatment and control conditions. Both groups cite the same criteria in support of their models, even though the control condition did not develop public criteria for model-goodness.

Table 1: Justifications given for students' self-assessment of model.

Criteria Referenced by Students in Their Evaluation	Number of Responses (%)	
	Control	Treatment
I explained my answer	14 (28%)	15 (38%)
I used the evidence	15 (30%)	16 (40%)
I drew a picture/diagram	26 (52%)	24 (60%)
I used detail	4 (8%)	1 (3%)
I used labels	7 (14%)	9 (23%)

While none of the students cited their class criteria about mechanism as a commendable aspect of the model (relevant criteria from their class lists were: "shows logical process", "shows sequential steps", "has mechanism"), several students did cite a lack of mechanism as a problem with their model. There were 13 students who did not evaluate their model positively. Of these, eight argued that their model was not great because, " I did not explain my reasoning on what I think is happening inside the bodies of people with DEB", "I am very unsure with how the blisters are formed", "I did not say how it creates blisters." Hence, for these students, the criteria lists helped to highlight a critical gap or shortcoming of the model. The students were unable to address the gap but did acknowledge it, an important epistemic achievement in itself. That the remaining 79 students did not identify this shortcoming is troubling. Then again, since most of the class criteria focused on communicative features of the models it is not entirely surprising that students concluded that their models were adequate if they clearly portrayed the phenomenon and addressed the evidence.

Students' Arguments for the Best Model

We next analyzed students' written arguments in support of their chosen models (separatin or connectin explanations). Overall, 63% of the students chose the correct connectin model, 32% chose the separatin model, and the remaining 5% were undecided. In their arguments, 71% of the students cited at least one piece of evidence in support of their model choice. With the majority of those (49 students) citing two or three pieces of evidence. Twelve of the 90 students explicitly cited the quantity of evidence supporting the connectin model as the reason they chose it. Students' use of evidence varied from merely noting a single piece of evidence, to discussing several pieces of evidence and explaining how the evidence supports the model (justifying evidence-model link). Table 2 illustrates students' use of evidence in their arguments. Note that categories are not mutually exclusive and students' arguments could be double coded. Overall, there seem to be no significant differences between the control and treatment condition, with one exception: twice as many students in the control provided explanations of how the evidence supports the model. We are not sure why this is the case. One possible explanation is the focus on fit-with-evidence and justification was, for some reason, made more salient in the control classes. We currently do not have evidence to support or refute this conjecture.

Table 2: Students Use of Evidence

Nature of Evidence Citation	Number of Responses (%)	
	Control	Treatment
Student does not discuss any piece of evidence	15 (30%)	12 (30%)
Student cites at least one piece of evidence	36 (72%)	28 (70%)
Student explains how the evidence supports the model	21 (42%)	10 (25%)
Student discusses how evidence relates to the competing model	9 (18%)	9 (23%)
Student mentions the quality and/or relevance of the evidence	3 (6%)	4 (10%)
Student notes amount of evidence as contributing to the choice	5 (10%)	7 (18%)

Across both conditions, there were seven students who noted more than one piece of evidence, provided a justification linking the evidence pieces to the model, and considered the competing model in their argument — a fully articulate argument and rebuttal:

> I think the connectin explanation is better. In Evidence 2 it says that when scientists studied normal and affected skin samples. The affected skin had large gaps between the dermis and epidermis just like explanation 1 says. In evidence 5 it says that they compared the proteins and noticed a protein was missing. In this explanation it's saying the connectin protein is missing. In evidence 6 people with DEB are injected with the connectin protein and their skin

> became better. If they had separatin protein it would just break the protein again. This shows that they just never created the protein. That is why the connectin explanation is better.

This student mentions three pieces of evidence (2, 5 and 6) that all support the chosen, and correct, model. The student further explains how the evidence pieces support the model. These evidence pieces point to a protein being missing, which is the core difference between the two models. The student understands this key distinction and notes that evidence 6 thus contradicts the separatin model ("If they had separatin protein it would just break the protein again"). This is a well-articulated, evidence-based, and justified argument and counterargument. Interestingly, five of the 29 students who chose the incorrect model (separatin) were also able to provide arguments that used evidence to refute the, actually correct, model:

> I think the separatin protein explanation is better: (1) Evidence 5 shows how people affected are missing a protein, like in the model; (2) Evidence 6 agrees with the other model, but is only done with 10 people; (3) it [separatin model] is clearer to me and makes way more logical sense.

This student interprets evidence 5 as supporting the incorrect model and then essentially dismisses the conflicting evidence (6) due to sample size. This student, who was in the treatment condition, was one of two students who discussed evidence quality, such as sample size, in their argument. The misinterpretation of evidence 5 was more common (24 students) and rather interesting.

Evidence 5 stated: "Scientists compared the different types of proteins found in the skins of normal and affected people. Affected people were missing one type of protein". Our intent with this evidence was to support the connectin model, which stated that affected individuals are missing the needed connectin protein. Many students did interpret the evidence accordingly:

> Evidence 5 contradicts explanation #1 because evidence 5 says that they are missing one type of protein but explanation one says that they get a *new* protein, which means that have an extra one, they are not missing one.

However, since the separatin model noted that the new separatin protein breaks down the connective protein that holds the layers together, several students interpreted the evidence in accordance with that model:

> Evidence 5 states that a protein is missing. This supports explanation #1 [separatin] because of how explanation 1 stresses that the separatin protein breaks down the protein which means it's not there therefore it supports explanation #1 because the protein is broken down.

While the first interpretation is correct, the second interpretation is sensible. In hindsight, it is clear that this evidence was problematic. Yet, this situation also highlights a design challenge: identifying evidence for evidence-based tasks that is neither too straightforward and simple, nor too ambiguous and open to multiple interpretations. On the one hand, scientific evidence can often be interpreted in multiple ways and understanding this point is an important epistemic achievement. Helping students learn that models are often under-determined by evidence can, and should, be an instructional goal that is facilitated by well-designed materials that present students with ambiguous and controversial evidence. On the other hand, there are specific content goals associated with curriculum materials and in these model-evaluation tasks it is necessary to craft the body of evidence to support the scientifically normative model. In this case, the ambiguity allowed a substantial number of students to choose the erroneous model and support it with evidence. We have no simple solution to this tradeoff in design, but it is a relevant and recurring tradeoff that requires careful consideration.

Conclusion and Implications

Returning to our research questions: (a) can students generate mechanistic models of the cellular and molecular mechanisms that underlie genetic phenomena? And (b) in what ways do students use evidence and epistemic criteria in their arguments? Our findings suggest that overall students were not able to generate mechanistic explanations of the cellular and molecular basis of the genetic phenomenon described in the assessment. Their models were mostly at the phenomenological level and essentially reiterated the symptoms. Evidently, the curriculum as designed and enacted was insufficient in helping students develop a more generalized schema of genetic mechanisms that they could apply to novel contexts. This finding underscores a core implication- that it is essential to help students develop generalized models/schemas of mechanisms in the discipline, and that multiple examples are likely needed to support the development of such generalizations.

In terms of the second research question, it seems that most students demonstrated awareness of some core epistemic criteria for good models and arguments. In evaluating their own models, most students referred

to secondary epistemic criteria related to communicative features of model (labels, clarity, pictures, etc.), and about a third mentioned the primary epistemic criteria of fit-with-evidence. This was the case for both study conditions, which is somewhat surprising since there was no explicit focus on criteria in the control condition. However, as we have shown in prior research (Pluta, et al., 2011), students are capable of coming up with both secondary and primary epistemic criteria for good models without much scaffolding. While students in the control condition did not develop public criteria lists, it seems that the constant discussion of evidence and models likely helped them develop a set of implicit criteria that they used in this task. In contrast, students in the treatment condition did not seem to develop substantially more sophisticated criteria. We wish to caution, however, against over generalization of these initial findings. The work reported here is based on assessments from one of four teachers and on a single task. It may be that this task context afforded less opportunity for students to demonstrate their developing epistemic prowess. In a prior study with middle school teachers from a different school, who had more experience with model-based inquiry and modeling criteria, we have shown that a focus on a related set of criteria (criteria for good evidence and criteria for determining evidence-model relationships) did result in significant gains in argumentation. Students' arguments included more explicit justification of how evidence related to the model, discussed quality of the evidence more often and in more detail, and used more evidence, including counterevidence for the competing model (Dianovsky, et al., 2013). Others have also had similar success in focusing student attention on epistemic criteria (Schwarz & White, 2005). We suspect that while the teachers in the study did engage students in the development, revision, and use of criteria, these criteria lists remained rather intuitive and superficial. The implication here is that to reap the benefits of explicit and public engagement with epistemic criteria, it is important to move beyond what students can do on their own and to really deepen, expand, and enhance their initial lists. In particular, progressing beyond communicative criteria to those that deal with more subtle aspects of model-evidence-fit, explanatory nature of models, accuracy, etc. appears to be essential.

Our findings also suggest that students were able to use evidence, provide reasons and justify their claims, at least to some extent. Almost half of the students in our study used multiple pieces of evidence to support their claims, a third provided justifications that explain how the evidence related to the model, and almost a quarter provided counterarguments against the competing model. These findings echo and extend research of others who have shown that, with proper support, students can use and internalize argument structure, and develop better evidence-based arguments (Duschl, 2007; McNeill, Lizotte, Krajcik, & Marx, 2006; Osborne, Erduran, & Simon, 2004).

Lastly, we wish to discuss some design implications for evidence choice in tasks that require students to develop or evaluate models. In the first task of the assessment, most students evaluated their models as being good because they addressed all the evidence, which was the case. There were no pieces of evidence that compelled students to provide mechanistic explanations at the molecular level. This was by design; we wanted to see if students would be able to come up with such mechanisms on their own and based on what they had learned in the unit. This expectation may have been too ambitious and we may have seen better models had one of the evidence pieces mentioned a protein or a genetic mutation, thereby cuing students to think about the molecular entities they had studied. The implication we draw here is that it may be necessary to provide evidence that relates to the explanatory mechanism one wants students to generate. This may seem fairly obvious, however, there is a tradeoff between providing evidence that "gives away" the answer, and providing evidence that directs students towards the appropriate grain size and nature of the desired explanation. A related tradeoff was also evident in the design of evidence 5 in the second part of the task. The evidence was somewhat ambiguous and open to multiple interpretations and students capitalized on this property and used the evidence in support of the incorrect model. Thus, there is also a delicate balance between choosing evidence that is too clear cut and that hides the under-determined nature of most real-world evidence-model relationships, and providing evidence that invites alternative interpretation and does not clearly rule out the erroneous model. We do not have guidelines or solutions to address these tradeoffs, but we believe the field would benefit from more explicit discussion of the design challenges intrinsic to engaging students with authentic disciplinary practices.

References

Chinn, C.A., and L.A. Buckland. 2012. Model-based instruction: Fostering change in evolutionary conceptions and in epistemic practices. In Evolution challenges: Integrating research and practice in teaching and learning about evolution, ed. K.S. Rosengren et al., 211–232. New York: Oxford University Press.

Dianovky, M., Duncan, R.G., & Chinn, C. A. (2013). Using Evidence to evaluate multiple competing models. Paper presented at the annual meeting of the *National Association of Research in Science Teaching.* Rio Grande, Puerto Rico.

Doerr, H., & Lesh, R. (2003). Foundations of a model and modeling perspective on mathematics teaching, learning, and problem solving. *In Beyond Constructivism*, pp. 3-33.

Driver, R., & Leach, J. M. R., Scott, P.(1996). *Young people's images of science*. Open University Press, Buckingham

Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. Journal of Research in Science Teaching, 44(7), 938-959.
Duncan, R. G., & Tseng, K. A. (2011). Designing project-based instruction to foster generative and mechanistic understandings in genetics. *Science Education,* 95(1), 21-56.
Duncan, R. G., Freidenreich H. B., Chinn C. A., & Bausch, A. (2011). Promoting middle school students' understanding of molecular genetics. *Research in Science Education,* 41(2), 147-167.
Duschl, R. (2007). Quality argumentation and epistemic criteria. In S. Erduran, & M. Jiménez-Aleixandre (eds.), *Argumentation in science education: perspectives from classroom-based research* (pp. 159-175). Amsterdam: Springer.
Giere, R.N. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742–752.
Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21, 725–740.
Grosslight, L., Unger, C., Jay, E., & Smith, C.L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28, 799–822.
Kitcher, P. (1993). *The advancement of science*. Oxford: Oxford University Press.
Kuhn, T.S. (1977). *The essential tension*. Chicago: University of Chicago Press.
Latour, B. (1987). *Science in Action*, Cambridge, MA: Harvard University Press.
Lehrer, R., & Schauble. L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology: Vol 5*. *Educational design and cognitive science* (pp. 101-159). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: re-visiting students' understandings of genetics. *International Journal of Science Education*, 26(2), 195-206.
Lewis, J., &Wood-Robinson, C. (2000). Genes, chromosomes, cell division, and inheritance- do students see any relationship? *International Journal of Science Education*, 22, 177–195.
Longino, H. E. (2002). *The fate of knowledge*. Princeton, NJ: Princeton University Press.
Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, 35(4), 183–189.
Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic phenomena. *Journal of Biological Education*, 34(4), 200–210.
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.
National Research Council (1996). *National science education standards*. Washington, DC: National Academies Press.
National Research Council (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.
NGSS Lead States. (2013). *Next Generation Science Standards: For States, By State*s. Washington, DC: The National Academies Press.
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41(10)*, 994–1020.
Pluta, W. J., Chinn, C. A., Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, *48*, 486-511.
Schwarz, C., & White, B. (2005). Meta-modeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23*(2), 165-205.
Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., … & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners.*Journal of Research in Science Teaching*, *46*(6), 632-654.
Treagust, D. F., Chittleborough, G. D., & Mamiala, L. T. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education,* 24, 357-368.
van Mil, M. H., Boerwinkel, D. J., & Waarlo, A. J. (2013). Modelling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behaviour. Science & Education, 22(1), 93-118.
Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretive framework. Journal of Research in Science Teaching, 35, 1031–1055.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1008634. 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.

Author Index

Pages 1-622: Volume 1
Pages 623-1176: Volume 2
Pages 1177-1764: Volume 3

Abbott, Robert D., 962, 1152
Abrahamson, Dor, 23, 1593
Acosta, Alisa, 673
Adams, Deanne, 1199
Aguilar, Stephen, 1665
Ahn, June, 174, 455, 657, 1719
Alcalá, Lucía, 13
Aleven, Vincent, 977, 1352
Alfonso-Gurneau, Jasmine, 12
Alibali, Martha W., 479, 649, 1042
Allert, Heidrun, 238
Alonzo, Alicia, 1037
Alston, Alice, 410
Alvarenga, Claire, 1012
Anderson, Emma, 118, 1456
Anderson, Janice, 1641
Andrade-Lotero, Alejandro, 1637
Arastoopour, Golnaz, 150, 1680
Arias, Anna Maria, 1426, 1749
Arici, Anna, 697
Asterhan, Christa S. C., 1342, 1684
Azevedo, Roger, 309, 1052
Bachfischer, Agnieszka, 1283
Baker, Ryan S., 222
Ballweber, Christy, 1647
Banerjee, Amartya, 1603
Bang, Megan, 4, 12, 1372, 1436
Bannister, Nicole, 1209
Barab, Sasha, 697
Barany, Amanda, 1199
Barber, Jacqueline, 1117
Barker, Lisa M., 1446
Barrientos, Kristina, 1012
Barron, Brigid, 1264
Barth, Armin, 1179
Barth-Cohen, Lauren, 325, 1531
Barzilai, Sarit, 721
Basu, Satabdi, 1097
Baumeister, Antonia E. E., 38
Beauvineau, Yves, 1022
Beck, Luisa, 1623
Bell, Alexander, 1082
Bell, Philip, 1228, 1426, 1710
Bemis, Carrie Allen, 1489
Ben-David Kolikant, Yifat, 1362
Ben-Zvi, Dani, 394, 1549, 1677
Bernstein, Debra, 1485
Berson, Eric, 1537
Bevan, Bronwyn, 1711
Bhatnagar, Sameer, 982
Bielaczyc, Katerine, 1315, 1677
Biemans, Harm J. A., 1569
Bientzle, Martina, 102
Birmingham, Daniel, 952
Biswas, Gautam, 1097, 1352
Black, John B., 230
Blair, Kristen P., 1179
Blikstein, Paulo, 863, 1147, 1669
Bolling, Amy, 1436
Bolzer, Markus, 1416
Boncoddo, Rebecca, 479, 649
Bonsignore, Elizabeth, 174, 455, 657
Booker, Angela N., 919
Borge, Marcela, 753
Boston, Melissa D., 997
Bouchet, François, 309, 1052
Bowker, Geoffrey C., 6
Boxerman, Jonathan, 1583
Brady, Corey, 1199, 1388, 1603
Bransford, John, 1647
Brennan, Karen, 18, 1559
Breuleux, Alain, 14
Briseño, Adriana, 879
Britt, M. Anne, 1541
Brodie, Karin,
Brooks, Christopher, 1691
Brown, Willard, 1571
Bryant, Julie, 1643
Buckingham Shum, Simon, 150, 1680
Burke, Jeff, 1436
Burke, Quinn, 86, 1219
Burkett, Candice, 1541
Burleson, Winslow, 278, 847
Buxton, Cory, 1332
Caccamise, Donna, 1002
Cadeiras, Martin, 1012
Caires, Roxane, 495
Cakir, Murat Perit, 1112
Calabrese Barton, Angela, 952
Caleon, Imelda, 535
Callanan, Maureen, 1228
Cantarero, Andrea, 1563
Capps, Daniel, 325, 1531
Carlone, Heidi, 1332
Carney, Michael, 1456
Cartier, Jennifer, 599, 1621
Cartun, Ashley, 348
Castillo, Tim, 1563
Cerratto Pargman, Teresa, 1597
Cervantes, Francisco, 1559
Chae, Hui Soo, 1709
Chaffee, Rachel, 1557

Champney, Danielle, 62
Chan, Carol K. K., 126, 333
Chao, Jie, 1523
Charles, Elizabeth S., 982
Chase, Kiera, 23
Chen, Gaowei, 583
Chen, Vivian, 1669
Chen, Xiaodi, 1623
Chen, Ying-Chih, 641
Cheng, Britte Haugan, 1661
Cheng, Harry, 1609
Cheng, Julius, 1645
Chi, Michelene T. H., 847, 972, 1527
Chi, Min, 1645
Chin, Doris B., 1179
Ching, Cynthia Carter, 1273
Chinn, Clark A., 615, 1122, 1189, 1686
Chiu, Jennifer L., 1523
Cho, Young Hoan, 535
Choi, Gi Woong, 1067
Choi, Jinnie, 607
Choi, Sung-Youn, 1635
Chowning, Jeanne, 1426
Chu, Haiwen, 1717
Chung, Huy Q., 418
Clariana, Roy B., 1543
Clark, Douglas B., 1199, 1342, 1388, 1657
Clarke, Sherice N., 583, 1684
Clarke-Midura, Jody, 1731
Clase, Kari L., 1567
Clegg, Tamara, 174, 455
Clement, John J., 503
Clodfelter, Erika, 1052
Close, Eleanor W., 1533
Close, Hunter G., 1533
Cobb, Paul, 4
Cober, Rebecca, 14, 1273
Coffey, Janet E., 1406
Cole, Michael, 1254
Collier, Wesley, 1680
Collins, Allan, 1315
Collins, Jamie, 1625
Conforti Preszler, Noelle, 1647
Conlin, Luke D., 31
Conn, Jessica, 1533
Conner, Laura, 1555
Constantin, Ana-Maria, 1511
Cook, Melissa Sunshine, 625
Cooper, Stephen, 992
Coppens, Andrew, 13
Cordy, Michelle, 1521
Core, Mark, 1057
Corona Caraveo, Yolanda, 13
Correa-Chávez, Maricela, 13
Cox, Christopher, 1509
Cress, Ulrike, 102
Crooks, Noelle, 1042
Cruz, Daniel, 1012
Cunningham, Christine M., 1587
Cunningham, Jahneille, 1671
Curnow, Joe, 134, 206
Cutler, Christopher T., 839
D'Amico, Laura, 1362
D'Angelo, Cynthia, 1489, 1732
Dalvi, Tejaswini, 1565
Daly, Alan J., 426
Damşa, Crina, 440, 1283, 1733
Danielak, Brian A., 1047
Danish, Joshua A., 1273, 1323, 1637
Dasgupta, Chandan, 1497
Davenport, Jodi L., 1583
Davis, Elizabeth A., 1426
Dayton, Andy, 13
DeBarger, Angela Haydel, 1022, 1703
DeJaegher, Crystal J., 1523
DeLiema, David, 1750
DeSutter, Dane, 987, 1599
Dede, Chris, 1579, 1581
Deitrick, Elise, 591
Delen, Ibrahim, 947
Delgado, Cesar, 262
Dempsey, Mary, 12
Deng, Mario C., 1012
Denner, Jill, 1007
Derry, Sharon J., 370, 1315
DiGiacomo, Daniela K., 70, 729
DiSalvo, Betsy, 793
DiSalvo, Carl, 793
Dianovsky, Michael T., 816
Dillenbourg, Pierre, 15, 1017
Dingyloudi, Filitsa, 761, 1132
Dixon, Colin, 1591
Doane, William E. J., 1047
Dookie, Lesley, 402, 1751
Dow, Steven, 1515
Drake, Joel, 1617, 1659
Duarte Olson, Izabel, 1372
Duck, Jennifer, 947
Duckles, Joyce, 737
Dugdale, Michael, 982
Dukeman, Anton, 1097
Duncan Valentine, Keri, 745
Durik, Amanda M., 713
Duschl, Richard,
Dussault, Mary, 1511
Dutilly, Erik, 1302, 1527
Easterday, Matthew W., 317
Eberbach, Catherine, 1228
Eberle, Julia, 463, 1734
Edelson, Daniel C., 1466
Edwards, Ann R., 1406
Eilam, Billie, 937
Eisenberg, Michael, 190
Eitel, Karla Bradley, 1509
El Taraboulsi, Sherine, 1302
Elby, Andrew, 286, 1037, 1406
Elinich, Karen, 1219, 1456
Engelmann, Katharina, 246
Engelmann, Tanja, 1543, 1545
Engeström, Yrjö, 1254, 1308