Learners’ Epistemic Criteria for Good Scientific Models
William J. Pluta, Clark A. Chinn, Ravit Golan Duncan

Graduate School of Education, Rutgers University, 10 Seminary Place,
New Brunswick, New Jersey 08901

Received 27 May 2010; Accepted 28 February 2011

Abstract: Epistemic criteria are the standards used to evaluate scientific products (e.g., models, evidence, arguments). In this study, we analyzed epistemic criteria for good models generated by 324 middle-school students. After evaluating a range of scientific models, but before extensive instruction or experience with model-based reasoning practices, students generated lists of criteria of good scientific models. Students’ individual lists of criteria were compared to expert criteria, identified by philosophers of science, and with findings from previous research on students’ understanding of modeling. The most commonly listed criteria referred to the clarity, pictorial form, and explanatory function of models. Almost a quarter of the students included criteria relating to model fit with evidence. Students’ criteria provided insights into their understanding of the explanatory and descriptive goals of models; the constitutive, communicative, and epistemic features of models; and the role of evidence in supporting models. Collectively, students demonstrated familiarity with a wide range of modeling ideas that can be leveraged in instruction to promote deeper understandings of the modeling practice. We argue that inquiry-based science instruction should include a strong emphasis on epistemic criteria.

Keywords: modeling; epistemic criteria; nature of science; reasoning

Epistemic criteria play an important role in science; they are the standards scientists use to evaluate the validity and accuracy of scientific products such as models, arguments, and evidence (e.g., Kuhn, 1977; Laudan et al., 1986; Longino, 2002). For example, scientists use criteria for judging the quality of scientific models and guiding the choice between alternative models (Kuhn, 1977; Newton-Smith, 1981; Sober, 1988); some of these criteria include:

- Good models have high levels of conceptual coherence and clarity.
- Good models are compatible with theories in other fields.
- Good models are appropriately parsimonious.
- Good models are consistent with empirical evidence.
- Good models have a history of making novel empirical predictions.

An example of an epistemic criterion for evaluating evidence is to prefer robust, readily replicable evidence that has large effect sizes (Staley, 2004). Arguments that explicitly or tacitly refer to these and other epistemic criteria are common in the written and oral discourse of scientists (see, e.g., Bazerman, 1988; Staley, 2004).

Despite the importance of criteria in everyday scientific practice, there has been minimal research addressing science students’ tacit or explicit knowledge, understanding, and use of epistemic criteria (Chinn, Buckland, & Samarapungavan, 2010, in press). The purpose of this study is to address this gap by examining middle-school students’ explicit ideas about epistemic (and nonepistemic) criteria for model goodness.
We focused on criteria for model goodness because of the recent emphasis on modeling in both the history and philosophy of science and in science education. In the history and philosophy of science, recent work has emphasized the central roles that the development, testing, evaluation, and refinement of models play in scientific practice (e.g., Giere, 2004; Godfrey-Smith, 2006; Longino, 2002; Nersessian, 2002; Suarez, 2003). In science education, model-based instruction has been a focus of many research programs (e.g., Chinn, Duschl, Duncan, Buckland, & Pluta, 2008; Raghavan & Glaser, 1995; Schwarz et al., 2009; Stewart, Hafner, Johnson, & Finkel, 1992; White, 1993; Windschitl, Thompson, & Braaten, 2008). As a result of this research, recent recommendations for reforming instruction have advocated placing modeling activities at the center of the curriculum (National Research Council, 2007). A major challenge in any inquiry program, including model-based programs, is developing methods to move students’ reasoning forward. An important dimension of growth in reasoning is students’ understanding and use of epistemic criteria. Therefore, understanding students’ ideas about criteria for model goodness is an important step in developing better methods for promoting reasoning.

There are several reasons for an increased focus on epistemic criteria in science classes. First, the use of epistemic criteria is a significant scientific practice in its own right. Understanding criteria and criteria-related practices is an important part of learning how to participate in science, as well as understanding the nature of science (NOS). The use of criteria is embedded in the modeling, argumentation, and evidence evaluation practices that are central to inquiry curricula. Second, a focus on criteria provides a fruitful instructional target for scaffolding. Making epistemic criteria explicit in classrooms is a good way to make the cognitive practices of science visible. Students can, for example, be encouraged to develop and revise their own criteria for model goodness, argument strength, and so on. Reflection on criteria can help students gain mastery of new and challenging inquiry practices. Finally, the social processes of debating, vetting, adopting, and applying epistemic criteria in group and class discussions can contribute to the development of powerful knowledge-building communities within classrooms (cf. Brown & Campione, 1996; Scardamalia & Bereiter, 2006). Public discussion about criteria can enable classes to develop norms that support better group and individual learning.

In this article, we report on a study of middle-school students’ naïve (preinstructional) ideas about epistemic criteria for model goodness. We report individual students’ lists of criteria generated early in the school year, before the students had extensive experience with modeling and before they had discussed epistemic criteria as a class. Students evaluated models before generating criteria in order to provide a sense of what is meant by “scientific model” and to activate preexisting knowledge about scientific models. This research addresses the following questions: (a) What are middle school students’ initial ideas about modeling and the quality of models before they engage in intensive modeling practice or extensive work with criteria? (b) To what extent do student-generated epistemic criteria match the criteria used by practicing scientists? (c) What does the overall sophistication and diversity of students’ initial criteria suggest about appropriate and effective instructional approaches and strategies that can promote growth in understanding of scientific models and modeling criteria? Our analyses provide information about students’ understanding of modeling and science, leading to important insights into how to develop instruction that can help students advance beyond their naïve ideas about epistemic criteria.

Before describing the study, we review relevant literatures, including (a) models and epistemic criteria in science, (b) model-based instruction and epistemic criteria, (c) research on students’ use and understanding of epistemic criteria, and (d) students’ epistemic criteria and their conceptions of models. We conclude with a discussion of how an understanding of students’ naïve epistemic criteria can provide insights into the development of better inquiry-oriented instruction.

Models and Epistemic Criteria in Science

Like many science education researchers (Hogan & Maglienti, 2001; Penner, Giles, Lehrer, & Schauble, 1997; Windschitl et al., 2008; see also National Research Council, 2007) we draw on philosophers’ work in conceptualizing scientific models. Models are idealizations that scientists use to represent aspects of the world for specified purposes (Giere, 2004, p. 742). Scientific hypotheses make claims about similarities between models and real systems (Giere, 1988, p. 81), claims that can be empirically tested (Longino, 2002; Solomon, 2001). Well-known models include representations of evolutionary mechanisms, including genetic
drift, natural selection, and transmutation; models representing the orbits of planets, such as the Copernican heliocentric and Ptolemaic geocentric models; and models depicting atoms such as the plum-pudding and Bohr’s models. Much of the work of scientists centers on developing models and investigating the extent to which models resemble the world in the intended respects (Giere, 1988; Kitcher, 1993; Nersessian, 2002; Niiniluoto, 2002).

Some of the model-specific cognitive practices performed by scientists include developing and revising explanatory models in response to evidence, evaluating the internal and external consistency of models, choosing among models, conducting experiments and other studies to test models, and evaluating the quality and strength of evidence. Social activities are essential to carrying out these cognitive practices; these include distributing and coordinating the work of experimentation and observation, communicating findings within the scientific community, and engaging in argumentation and co-construction of knowledge. Discourse is central; scientists critique each others’ models, provide support and elaboration to further develop models, and appropriately reconcile alternate models, approaches, and ideas. Many of these strategies and practices are incorporated into model-based learning environments.

The use of epistemic criteria or standards is central to many of the practices outlined above. To decide whether a model is adequate, or whether one model is better than another, scientists and science students alike must evaluate models utilizing standards or criteria of model goodness. As we noted earlier, these criteria can include conceptual coherence, evidential fit (e.g., scope of evidence covered, degree of fit with that evidence), and parsimony. Similarly, when judging the supportive strength or fit between models and evidence, reasoners must evaluate the strength of evidence. Criteria for evaluating evidence strength include standards used to judge the methodological strengths and weaknesses of empirical studies. For instance, scientists may maintain that the strongest evidence derives from empirical studies that have appropriate controls, adequate sample size, use appropriate measures, and include multiple measures that give converging results. Epistemic criteria may be deployed with or without explicit awareness.

Criteria related to model-evidence fit and coherence with other well-established theories can be viewed as central to evaluating a scientific model. These primary epistemic criteria center on the likely accuracy of the model, by which we mean the extent to which the model resembles intended aspects of the world in desired respects, as indicated by fit with empirical observations. Other criteria—such as criteria for how clearly models are presented—are secondary criteria because they do not directly impact the accuracy of a model; an accurate but poorly presented model can be represented in a clear way without changing any substantive content. In the present study, we classify student-generated criteria in terms of whether they are primary epistemic criteria, which are central to the epistemic practices of science (identified by philosophers of science) because they focus on the accuracy of the model, or secondary epistemic criteria, which do not directly impact the accuracy of the model, but contribute to epistemic aims of science, such as communicating important ideas clearly so that others can use these ideas. Distinguishing between primary and secondary criteria provides a useful guide for evaluating students’ epistemic cognition and development during generation and use of criteria for good models. It provides a rough indicator of the extent to which students are focused on accuracy of the model versus other model criteria that are important, but can be met by inaccurate as well as accurate models. Instruction aimed at introducing or developing understanding and use of criteria can then be targeted at specific levels of epistemic criteria.

Model-Based Instruction and Epistemic Criteria for Good Models

There is promising evidence that model-based instruction improves students’ ability to construct their own models (White, 1993), develop explanations and conclusions from models (White & Frederiksen, 1998; Schwarz & White, 2005), competently coordinate and work with multiple models (Gutwill, Frederiksen, & White, 1999), synchronize models and evidence (Pluta, Buckland, Chinn, Duschl, & Duncan, 2008; Snir, Smith, & Raz, 2003; Zimmerman, Raghavan, & Sartoris, 2003), critique and revise models (Penner et al., 1997; Stewart et al., 1992), and engage in causal and analogical reasoning (Harrison & Treagust, 2000; Raghavan, Sartoris, & Glaser, 1998). Model-based inquiry can promote conceptual change, as well (Chinn & Samarapungavan, 2009). Overall, these studies indicate that students engage in impressive reasoning when given the opportunity to practice modeling.

*Journal of Research in Science Teaching*
As model-based inquiry environments become more prevalent, there is a need for research that investigates methods of scaffolding reasoning in these environments. Existing research emphasizes helping students make critical structural distinctions, such as the distinction between explanations and evidence (e.g., Bell & Linn, 2000; McNeill, Lizotte, Krajcik, & Marx, 2006; Suthers & Hundhausen, 2003; Toth, Suthers, & Lesgold, 2002). Other research explores scaffolds that support understanding of an inquiry cycle for modeling (e.g., Schwarz & White, 2005; Schwarz et al., 2009) and meta-awareness of core reasoning processes (White & Frederiksen, 1998). Scaffolding the development of epistemic criteria could be another very productive method of promoting growth in reasoning and understanding of the nature of science.

Scaffolds for epistemic criteria include methods that make student thinking about epistemic criteria public and encourage students to discuss criteria with peers and facilitators. One example of such a scaffold would be the public development of criteria for model goodness. Teachers could lead a class discussion in which students, having examined a variety of better and worse models, develop a list of public class criteria for model quality. In further class discussions throughout the year, students could periodically revise the class criteria as they gain more experience with evaluating models. They would use their own criteria to evaluate their own, peer, and teacher provided models throughout the year, and the criteria could also be the basis of rubrics used by the teacher to evaluate students’ work. We have assayed this approach in our work with model-based inquiry in middle schools (Chinn et al., 2008; Chinn, Pluta, Buckland, Rogat, DiFranco, & Witham, 2010).

In order to understand how to develop this and other scaffolds of epistemic criteria, it is valuable to understand students’ ideas about epistemic criteria. Instruction may profit from taking students’ naive ideas into account—much as knowing students’ prior conceptions about photosynthesis or about force and motion can guide the design of more effective instruction (Chinn & Brewer, 1993; Driver, Squire, & Wood-Robinson, 1994). At present, little is known about what students consider to be characteristics of good scientific models or how they think good models differ from poor models. With this information, we can develop instruction to help students learn progressively more sophisticated criteria.

Learning about epistemic criteria affords opportunities for students to learn about significant aspects of the nature of science. First, as students learn about epistemic criteria, they should learn that criteria cannot be applied rigidly or algorithmically; criteria must be applied contextually. For instance, it is often not straightforward to determine which of two or more alternative models has more supporting evidence, nor is it always obvious which of two theories is more parsimonious. Criteria can also conflict with each other (Laudan, 1977), as when a substantially more parsimonious model does not explain the evidence quite as well; different scientists can reasonably come to different conclusions about how to weigh competing criteria differently (Kuhn, 1977). Second, students can learn that rival scientists sometimes differ in favored epistemic criteria (Kuhn, 1977; Toulmin, 1958). For example, the history of biology has sometimes seen divergence in the criteria of some scientists who give precedence to experimental results and others who give greater significance to careful observations of nature (Mayr, 1982). This disagreement stems from a difference in criteria for what counts as the best evidence. Third, students can learn that epistemic criteria can change. As an example, consider the field of chemistry. Before Lavoisier, one core criterion for chemical explanations was that good models must explain color and texture, and changes in both. This criterion was much less central after Lavoisier. Over time, scientists critique the prevalent criteria; they may abandon or modify old criteria and develop new ones (Laudan, 1984; Longino, 2002). Like theories and models, criteria are accepted through surviving repeated rounds of social criticism. Criteria that do not survive this criticism may be abandoned, modified, or de-emphasized.

Research on Students’ Use and Understanding of Epistemic Criteria

There is an increasing amount of research on students’ ideas about criteria. Five prominent examples are studies by Samarapungavan (1992), Hogan and Maglienti (2001), Penner et al. (1997), Smith, Maclin, Houghton, and Hennessey (2000), and Schwarz and White (2005). Using a theory-choice task, Samarapungavan (1992) found that first, third, and fifth graders made theory choices that conformed to normative epistemic criteria recommended by philosophers of science, including range of explanation, non-ad hocness of explanation, and empirical and logical consistency. A majority of students could explicitly
justify their theory choice utilizing the criteria of explanatory scope (e.g., by saying that one theory explains more than the other) and of empirical consistency (e.g., by saying that one theory is not as good because some evidence goes against it). However, a majority of students did not explicitly articulate the normative criteria to justify correct choices that favored simpler theories or theories that avoided ad hoc explanations. In this study, students articulated reasons that drew on criteria, but they were not asked to reflect metacognitively or talk about the criteria themselves. Moreover, only those few criteria that were relevant to the forced theory choices were examined in this study; whether students would spontaneously propose these criteria (or other criteria) was not investigated.

Hogan and Maglienti (2001) compared students’, laypeople’s, and scientists’ reasoning about scientific conclusions, finding that students and laypeople appealed to personal inferences and values as a criterion for judging the validity of the conclusions, whereas scientists did so far less frequently. Scientists also appealed to criteria relating to precision and specificity in the conclusions; students did not. Students were not asked to describe or reflect on the criteria they used to evaluate conclusions.

Penner et al. (1997) found that first and second graders who had engaged in modeling instruction were more likely to evaluate models by appealing to criteria related to functional aspects of models, whereas students who had not engaged in modeling instruction were more likely to draw on criteria related to perceptual qualities of models. The researchers did not ask students to be explicit about their general criteria. Similarly, Smith et al. (2000) found that middle school students who engaged in constructivist science curriculum that included modeling generated more criteria for evaluating scientific beliefs than students who participated in more traditional science instruction.

In an instructional study, Schwarz and White (2005) had students evaluate models using four criteria that were directly taught to them: accuracy, plausible mechanism, consistency, and utility of models. Students evaluated their own, peer-generated, and researcher-generated models utilizing the provided criteria. Schwarz and White found that students who used model-criteria demonstrated a better understanding of the nature of modeling, inquiry, and physics content than students who completed the same instructional unit, but without explicit use of criteria (comparison data were drawn from a different study described in White & Frederiksen, 1998). This result provides evidence that instruction focused on epistemic criteria can scaffold science learning. However, this study does not provide information on whether students have the ability to generate and use their own criteria, or what students’ own ideas about criteria are.

Although these studies provide valuable information about students’ epistemic criteria for good models, none of them employed a method that seems promising in understanding students’ epistemic criteria—namely, asking students to reflect on their own criteria for what makes a model good. The study described in this article addresses this gap in prior research by asking students to report on their ideas about criteria met by good scientific models.

Students’ Epistemic Criteria and Conceptions of Models

Students’ conceptions about epistemic criteria are likely to be interconnected with their understanding of models. For example, if students hold the conception that models are literal copies of nature, they will likely fail to understand why models needs to be revised in light of evidence and, more specifically, will be unable to apply criteria related to the nuances of model-evidence fit. Similarly, if students fail to understand the idea that models frequently present mechanisms, then criteria such as “has a good explanatory mechanism” will be meaningless to them. This raises the question of whether students have a good enough understanding of models to even begin considering criteria for good models.

There have been only a few studies of students’ (or teachers’) understanding of models and modeling (Grosslight, Unger, Jay, & Smith, 1991; Schwarz & White, 2005; Treagust, Chittleborough, & Mamiala, 2002; Windschitl et al., 2008). Most research suggests that students’ views of models and modeling are inaccurate without instruction (e.g., Abd-El-Khalick, Bell, & Lederman, 1998; Carey & Smith, 1993; Grosslight et al., 1991; Treagust et al., 2002). For example, Lederman, Abd-El-Khalick, Bell, and Schwartz (2002) have found that few people recognize that scientists’ evaluations of models are tentative and somewhat subjective. Grosslight et al. (1991) found that in the absence of instruction, most secondary-school students thought of models as little more than direct replications or copies of reality (67% of mixed-ability seventh graders; 23% honors-level eleventh graders). Few understood that models are created for specific purposes.
(12% of seventh graders; 36% of eleventh graders), understood the role of evidence in developing models (0% of seventh graders; 45% of eleventh graders), or realized that models can generate predictions (Grosslight et al., 1991). Tregast et al. (2002) reported similar results, with two major differences. Tregast et al. (2002) found that over 70% of the students (ages 13–15) in their sample believed that models are revised in light of evidence, and nearly 50% agreed that models are used for “making predictions, formulating theories and showing how information is used” (p. 365). They also found that a preponderance of students viewed models as direct copies of nature.

Given the interrelatedness of epistemic criteria and modeling, eliciting students’ epistemic criteria may provide more information about students’ understanding of modeling. Although the current study does not attempt to be an inclusive survey of students’ understanding of modeling, it may potentially shed light on the reasons for some discrepancies in previous research. In addition, research on students’ understanding of formal science does not provide much insight into the epistemic knowledge students actually use during instruction (Louca, Elby, Hammer, & Kagey, 2004; Sandoval, 2005). Student-generated lists of criteria may more closely align with the knowledge students actually use and display during learning and inquiry.

Overview and Goals of the Study

The present study investigates criteria that students explicitly generate after examining and evaluating the quality of a range of scientific models. The present study is the first to explore students’ own generated sets of epistemic criteria. Generating these lists required metacognitive reflection on the criteria themselves. Students were presented with multiple models of varying purposes and qualities, which they discussed in pairs. Then individual students generated their own lists of criteria for what made good models.

These data potentially provide important information about students’ epistemic cognition. If students can generate reasonable sets of criteria on their own, then it will be possible for students’ criteria to become the basis for classes to construct community norms to use epistemic and other criteria for model goodness. Teachers can lead discussions in which students propose and critique various criteria for model goodness, culminating in an agreed-upon set of shared criteria. This would be a highly student-centered approach to developing shared norms regarding criteria. Before attempting to engage students in this type of discussion, a teacher would likely want to know the extent to which students collectively understand primary and secondary criteria and the extent to which these ideas are shared by students within a class.

For example, if the present study shows that students generate a combination of primary, secondary, and nonnormative criteria, the conversation could be geared towards the dialogic vetting of criteria for their effectiveness in choosing between alternate models; the teacher could perhaps focus discussion towards primary criteria or distinguishing between primary and other kinds of criteria. If students mostly generate primary epistemic criteria, then teachers could engage students in sharing their ideas, without needing to redirect the conversation toward criteria the students do not propose on their own. If, on the other hand, students cannot generate appropriate criteria, it suggests that more teacher-centered approaches may be needed to enable students to develop criteria. One of the aims of this study is to determine whether seventh-graders’ naïve understanding of criteria provides an adequate basis for a student-centered approach, or whether more teacher-centered approaches are needed.

In short, the central goal of this study was to identify students’ early, developing ideas about epistemic criteria for model quality. The primary source of data was student-generated lists of criteria, developed before students received extensive instruction on modeling, but after they had the opportunity to reflect on several scientific models similar to those found in textbooks. As mentioned previously, we addressed these questions: (a) What are middle school students’ initial ideas about modeling and the quality of models before they engage in intensive modeling practice or extensive work with criteria? (b) To what extent do student-generated epistemic criteria match the criteria used by practicing scientists? (c) What does the overall sophistication and diversity of students’ criteria imply about appropriate and effective instructional approaches and strategies?

Prior research provides grounds for both pessimism and optimism about the quality of students’ generated criteria for good models. On the pessimistic side, there is a large body of research that appears to demonstrate that students have a poor understanding of the nature of science (Abd-El-Khalick et al., 1998; Carey & Smith, 1993), of models in particular (Grosslight et al., 1991; Tregast et al., 2002), and of
theory-evidence relations (Zimmerman, 2007; cf. Bråten, Britt, Strømsø, & Rouet, 2011). These findings suggest that students’ criteria for model goodness might focus exclusively on issues of understandability or clear communication rather than on issues of evidential fit. This tendency could be reinforced by textbook models, which are intended to communicate scientific ideas rather than to be evaluated by students. Further, Schwarz and White (2005) found that students could use criteria that they were explicitly taught, but that most students could not reflectively articulate these criteria even after instruction. The ability to reflect on criteria requires a degree of meta-awareness of one’s epistemic cognition that might be beyond the reach of many seventh graders.

On the other hand, there are some reasons for optimism. First, Samarapungavan (1992) found that even elementary-school students can specify reasons implicating scope of evidence and consistency with evidence as reasons to prefer one theory over another. By the middle school age, students may have further developed the capacity to reflect metacognitively on their criteria. Second, previous research on models has not provided students with much experience with models before asking them questions about models; it could be that students do not understand the researchers to be referring to scientific models so much as to physical models, such as a model of a ship. If students have a better understanding of what the term model refers to, they may be better able to articulate criteria for model goodness. In this study, we guarded against misunderstanding of the task by exposing students to a range of scientific models and asking them to reflect on their quality before they developed their lists of criteria. We believe that this procedure provided students with a context for thinking about scientific models.

The study was part of a yearlong project involving seventh-grade life science students and teachers who were implementing a model-based inquiry program called PRACCIS (Promoting Reasoning and Conceptual Change in Science). This program of research has many features in common with other model-based inquiry programs (e.g., Raghavan, Sartoris, & Glaser, 1998; Sandoval & Reiser, 2004; Schwarz & White, 2005; Schwarz et al., 2009; Stewart et al., 1992; White & Frederiksen, 1998). This curriculum and research project (Chinn et al., 2008) aims to develop instructional schemes and tools which teachers can easily embed within their own instructional materials as well as to develop instructional modules lasting 1–4 weeks that teachers can use flexibly within the constraints of their own state and district curricula. A central instructional scheme in this project is the reasoning seminar. During reasoning seminars, students engage in argumentation in which evidence is used to construct, revise, and evaluate explanatory models.

The assessment of students’ own epistemic criteria for good scientific models reported in this article took place near the beginning of the year, before students had extensively engaged in modeling and model evaluation with criteria. Before developing their list of criteria, students reflected on a range of good and poor scientific models. This initial, orientating activity helped reduce confusion about what was meant by the term scientific model and provided an opportunity to reflect on the characteristics of good models. Thus, the assessment which followed this activity served as a measure of students’ initial ideas about models and criteria for good models.

**Method**

We examined lists of criteria generated by seventh grade students following a 40- to 50-minute activity aimed towards activating students’ existing knowledge of scientific models. In analyzing the results, we compare and contrast the criteria identified by students with the criteria identified by philosophers of science and with findings of previous research on students’ understanding of modeling.

**Participants**

Participants were 324 students in four seventh-grade teachers’ classes in two diverse, suburban New Jersey school districts (two teachers per district). The two districts reflected different levels of success on statewide tests of proficiency in mathematics and reading. In School 1, 1% of students qualified for free or reduced-price lunch; 97% of students were white. Students performed well on New Jersey state exams; approximately 90% of students in the school reached proficient or advanced proficient levels on state language arts and mathematics exams. In School 2, 27% of students qualified for free or reduced-price lunch; 47% of students were black, 27% were white, 15% were Hispanic, and 11% were Asian. The school’s performance on New Jersey state standardized tests was lower than in School 1. Approximately 70% of
students reached proficient or advanced proficient levels on the language arts exam; approximately 65% reached these levels on the mathematics exam. Students in 15 classes (at least three classes per teacher) participated in the study; each class was composed of students of mixed abilities.

Context

The task of generating individual criteria was embedded into each teacher’s regular instructional schedule. The assessment took place within the first month of school, immediately after parental consent and student assent were obtained. Students had not received significant instruction or experience with scientific modeling before the assessment. However, because teachers had received training and had adopted the model-based curriculum during the previous school year, some teachers used the word model or mentioned some modeling ideas to students before this activity. All four teachers had spent some time providing instruction on the scientific method (this instruction took place before the research team began collecting data; the research team did not facilitate any instruction during this period). This instruction was generally traditional instruction on topics such as measurement and experiments and did not focus on the core ideas that the modeling curriculum addressed—developing explanatory models to fit evidence, revising models in accordance with new evidence, incorporating mechanisms into models, considering the strength of evidence, understanding the nature of models, and so on.

Materials

Over a period of 40–50 minutes spanning all of one class period and part of a second, students completed a series of model evaluation tasks. We specifically aimed to: (a) provide a brief, working definition of models to students; (b) activate students’ existing knowledge of models, developed through viewing models in textbooks and other media; and (c) present contrasting cases that would help students begin reflecting on specific features of models related to quality.

Students were provided with a 13-page packet. The packet included the following explication of scientific models: “Scientific models are explanations. We can use scientific models to explain things in the world. Often, we can also use scientific models to predict things. For example, a model of how plants grow explains why and how plants grow, and it can help us make predictions about when plants will grow and when they will not grow.” We had two reasons for presenting this definition of models: (a) Developing explanatory and predictive models was a primary instructional goal for the larger research project. (b) Explanatory models, which can also be used for predictions, are a prevalent form of model in contemporary science; hence, learning to develop, revise, select, and use explanatory models is an authentic practice of scientists. Although there are other types of scientific models, philosophers of science (e.g., Giere, 1988; Kitcher, 1993; Longino, 2002; Machamer, Darden, & Craver, 2000) and science educators (e.g., Clement, 2000; Duschl, 2008; Schwarz et al., 2009; Stewart et al., 1992; Windschitl et al., 2008) have emphasized the centrality of explanatory models in science. Researchers such as Schwarz et al. (2009), Smith et al. (2000), and Windschitl et al. (2008) have focused on developing instructional schemes that promote students’ understanding and use of the explanatory models. Other researchers have examined data models (e.g., Lehrer & Schauble, 2004), emergent models (e.g., Penner, 2000), or focused on the analogical or cognitive aspects of models and modeling (e.g., Harrison & Treagust, 2000). These are important types of models in science, as well, but we restrict our focus to explanatory models in this article.

The first section of the packet displayed 12 different representations of volcanoes. Students circled the representations they thought were models and discussed their ideas with a partner. The representations included models, non-models, and representations that were debatable as to whether they were models. Non-models included pictures of actual volcanoes erupting and a volcano toy. We included both descriptive models (e.g., a static diagram of a volcano with only physical features labeled) and explanatory models. Four different model representations explaining how volcanoes erupt were included (flowchart, written explanation, causal diagram, and pictorial model with text). We also included a model explaining how dome volcanoes collapse, a formula for the force of a volcanic eruption, a data representation (a map displaying volcanic sulfur dioxide concentrations), and a schematic diagram of scale model. This activity aimed to provide grounds for reflecting on what kinds of representations are scientific models and what kinds are not scientific models, and to encourage students to think about scientific models they had previously.
viewed. Most students have seen scientific models in textbooks, television, or other media; however, they likely have not interacted with them in an active way (e.g., evaluating, choosing between alternative models, or revising them). Nor were students likely to have ever considered instances that are not models and thought about what distinguishes between models and non-models.

The second section of the packet presented seven pairs of models, each pair on a single page. The models addressed phenomena with which most students were somewhat familiar, including butterflies/lifecycles, global warming, food webs, amphibian lifecycles, the water cycle, plant growth, and diffusion. The two models in each model pair modeled the same general phenomena using two different representations. The dimensions along which the models differed (which correspond to the seven phenomena) included, (a) whether the models were descriptive or explanatory, (b) whether the models were explanatory or representations of data, (c) the degree of model complexity, (d) the presence and type of communicative features (e.g., labels, title), (e) the extent to which the models included details, including both relevant and extraneous details, (f) whether mechanisms were present and, if so, the types of mechanisms, and (g) the extent to which models were consistent with everyday data that students were likely to know about.

Each pair of models had one of two types of questions for students to consider. The first type of question was generic, “Which model is better? Or are they equally good? Or is it impossible to say which one is better?” The second type of question referred directly to a purpose for the model, such as, “Which of these two models is better if you want to explain how the smell of perfume can spread across a room?” These contrasting cases and the questions were designed to help students reflect on specific features of models that were related to model quality. We expected students to be relatively familiar with the phenomena modeled in this section and likely to be familiar with some of the specific representations.

Procedures

At the beginning of class, teachers informed students that they would be learning about scientific models. Teachers briefly discussed the explanation of models found on the first page of the packet. Students then worked through the introduction to modeling packet. For each page of the packet, students individually studied the representations and selected the best models. Students then discussed each page with a partner. As teachers monitored student progress, they encouraged students to explain their thinking to each other but generally sought to avoid providing specific information or feedback to students. Upon completing the packet, students individually generated and wrote six criteria for good models, following these instructions: “You have been thinking about the characteristics of good models and not-so-good models. Now, individually, make a list of the most important characteristics of good (rather than bad) models. Write down six important characteristics of good models that you can think of below.” In most classes, students completed the introduction to model packet in 1 day and generated their individual criteria on a second day. By the time students generated their individual criteria they were likely to have understood what we were referring to when we used the term scientific models. Students’ individual lists of criteria generated after the orientating activity are the sole focus of the analyses reported in this article.

Coding

We iteratively refined and developed codes through several analyses. In developing our coding scheme, we initially developed candidate categories based on criteria developed by philosophers, criteria suggested by the literature on students’ understanding of models (especially Grosslight et al., 1991 and Treagust et al., 2002), and an analysis of our introduction materials. Candidate criteria developed by philosophers included categories such as range of evidence explained, consistency with evidence (and avoiding inconsistency), testability, coherence, and simplicity. Candidate criteria suggested by the work on modeling included reference to kinds of models (e.g., explanatory and descriptive), use of different representational forms, and clarity. Our initial coding scheme included 47 categories. We revised the coding scheme by repeatedly examining student responses, adding to and modifying the initial candidate categories to better fit student responses. The initial categories were winnowed to 32 categories. Student responses assigned to each category were iteratively checked against each other to make sure that similar responses were consistently being coded in the same way.
Students often wrote the characteristics of good models as short phrases, such as “[a good model is] neat, detailed, descriptive” or “[a good model] shows how or why.” In such cases, we judged that the words in these phrases reflected different criteria. For example, in the first example, we coded the word neat as an instance of the criterion “good models are well organized,” which was defined as including neatness. The word detailed was classified as an instance of the category “good models are detailed.” And the word descriptive was coded as an instance of the criterion “good models are descriptive.” The second example, “shows how or why,” was coded as an instance of explanation criterion because an explanation has as a primary function giving an account of how or why things occur (Kitcher, 1981). All categories that emerged from the coding are described in the results section.

We coded all written lists of student-generated criteria. Intercoder agreement, computed based on 50% of the data coded by two coders, was 85%. Disagreements were resolved through discussion.

**Results and Discussion**

On average, each student generated 6.1 distinct criteria, and each class collectively generated a mean of 25.5 distinct criteria. Table 1 presents the percent of individuals providing each criterion, as well as the percent of classes in which at least one student generated each criterion. Overall, there were no major differences at the class, teacher, or school level in the distribution of individual criteria. We organized the criteria into five broad categories, which we discuss below. First, we discuss students’ criteria related to the goals of models, because understanding the goals of modeling is essential to understanding why other modeling criteria are important. Second, we discuss model constituents because these responses provide information about students’ basic ideas about what models should “look like” and thus how models are structured to accomplish their goals. We then consider communicative criteria, evidential criteria, and other epistemic elements in an order that roughly corresponds to the overall prevalence of responses within each category.

Within these categories, we further distinguished between three levels of criteria: (a) primary criteria, which are central to the practices of science (as observed by philosophers of science) and center on the likely accuracy of the model (by the terms accuracy, we refer to the extent to which the model is congruent with the world in desired respects, as indicated by congruence with empirical observations); (b) secondary criteria, which do not directly impact the accuracy of the model, but contribute to epistemic aims of science; and (c) criteria that are vague or suggest misconceptions about the practices of science.

In the following sections, we discuss individual and class criteria within each of the five categories of responses. In each section, we first present results and then discuss implications of our results. We also consider the implications for teachers who wish to lead class discussions in which classes propose and discuss criteria that can then serve as community norms for the evaluation of models; knowledge of the distribution of individual criteria in a class can influence how teachers go about facilitating these discussions.

**Goals of Models**

Criteria classified in the *Goals of Models* category reflect students’ ideas about the purposes and functions of models in scientific practice. As we noted earlier, a central aim of contemporary scientific research is to develop explanatory models—models that explicate mechanisms and identify causal and functional relationships. However, educational research suggests that students often view models as having the goal of literally depicting objects (Grosslight et al., 1991; Treagust et al., 2002). Students’ criteria relating to the goals of models have the potential to provide information on whether students see models as carbon-copy replicas of phenomena or whether they appreciate the broader goals of models.

In our sample, 90% of the students made some reference to the goals of models in their criteria. Thirty-seven percent of students noted only a single model goal, 33% noted two model goals, and 20% of students specified three or more criteria related to model goals on their individual criteria lists. Thus, 53% of the students in the study generated multiple goals for models in their lists of criteria. Individual students proposed an average of 1.5 criteria in this category. In the four most common responses, students wrote that models should provide *explanations* (55% of students), *information* (41% of students), *descriptions* (24% of students), and *answers to a question* (21% of students).

*Journal of Research in Science Teaching*
Table 1
Criteria for good model: student responses

<table>
<thead>
<tr>
<th>Criteria elements</th>
<th>Definition and example</th>
<th>Students (n = 324)</th>
<th>Classes (n = 15)</th>
<th>Interpretation of epistemic status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals of models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explains-1</td>
<td>Explicitly uses the word explain or explanation, e.g., the model should explain something</td>
<td>51% 100%</td>
<td></td>
<td>Primary. Philosophers frequently discuss the explanatory role of models (e.g., Giere, 1988; Machamer et al., 2000; Mayr, 1982). Because the introduction-to-models activity explicitly used the word explains, we distinguished between responses that used this word and responses that used other words.</td>
</tr>
<tr>
<td>Explains-2</td>
<td>Uses other descriptive language for explanation, e.g., shows how and why</td>
<td>4% 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Describes</td>
<td>Explicitly uses the word describe, e.g., a good model uses descriptive pictures.</td>
<td>24% 87%</td>
<td></td>
<td>Secondary. Although they do not address the central explanatory goals of science, purely descriptive models have played a role in the history of science (e.g., anatomy, botany), as well as in educational settings (Kuhn, 1977; Mayr, 1982).</td>
</tr>
<tr>
<td>Provides information</td>
<td>Suggests models consist of static or isolated pieces of information or facts, e.g., provides information</td>
<td>41% 100%</td>
<td></td>
<td>Vague. Although models do inform, this is not a central feature of models as discussed by philosophers. We think this language is too imprecise to take as evidence of a deep understanding of modeling.</td>
</tr>
<tr>
<td>Answer a question</td>
<td>Alludes to the idea that models provide answers to problems; responses usually use the words answer or question</td>
<td>21% 73%</td>
<td></td>
<td>Vague. It is unclear what students mean by this response. Different students might have different kinds of questions in mind.</td>
</tr>
<tr>
<td>Examples</td>
<td>Models are examples (models as illustrations of explanations), but not necessarily as abstract exemplars (an explanations), e.g., gives an example</td>
<td>8% 53%</td>
<td></td>
<td>Vague. This language is too imprecise to suggest a well-developed understanding of models. It is unclear whether students are referring to a concrete example or a more idealized example.</td>
</tr>
<tr>
<td>Data</td>
<td>Explicitly uses the word data, without suggesting that the data support the model</td>
<td>5% 53%</td>
<td></td>
<td>Secondary. Models of data are an important kind of model in science (Chinn &amp; Brewer, 2001; Staley, 2004). Students may be suggesting that models can present data or that models are built upon or supported by data.</td>
</tr>
<tr>
<td>Model constituents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures</td>
<td>Explicitly uses the word pictures</td>
<td>60% 100%</td>
<td></td>
<td>Secondary. Responses in the model constituents category specify features of models that are needed to help communicate what the models represent. Many views of science emphasize the importance of disseminating information to scientific peers (e.g., Bishop &amp; Trout, 2005; Goldman, 1999). If scientists cannot understand models, even if the models meet many other primary epistemic criteria, the models are valueless. Thus, although these criteria are not central criteria, they do have significant secondary value.</td>
</tr>
<tr>
<td>Words</td>
<td>References the use of words, texts, and language in models</td>
<td>30% 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagrams</td>
<td>Explicitly uses the word diagrams</td>
<td>19% 93%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labels</td>
<td>Explicitly uses the word label</td>
<td>13% 93%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrows</td>
<td>Explicitly uses the word arrows</td>
<td>12% 87%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visuals</td>
<td>Explicitly uses the word visuals</td>
<td>11% 73%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Explicitly uses the word title</td>
<td>7% 63%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing</td>
<td>Explicitly uses the word drawing</td>
<td>3% 33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicative elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity</td>
<td>Appeals to lucidity, e.g., model is easy to understand; model is clear</td>
<td>64% 100%</td>
<td></td>
<td>Secondary. As noted above, many views of science emphasize the importance of disseminating information to scientific peers. For these reasons, these criteria are conductive to the spread of scientific knowledge, and we classify them as secondary criteria.</td>
</tr>
<tr>
<td>Focus</td>
<td>Good models stay on topic</td>
<td>37% 93%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Refers to models being neat and organized. Arrows not all over the place</td>
<td>26% 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audience</td>
<td>References the audience that the models are designed for. Includes peers, scientists, or general public</td>
<td>5% 60%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Journal of Research in Science Teaching
### Table 1
(Continued)

<table>
<thead>
<tr>
<th>Criteria elements</th>
<th>Definition and example</th>
<th>Students (n = 324)</th>
<th>Classes (n = 15)</th>
<th>Interpretation of epistemic status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details</td>
<td>Appropriate details are represented in good models, e.g., just enough details</td>
<td>40%</td>
<td>100%</td>
<td>Primary. Ideas about appropriate details and complexity appear closely related to the normative criterion of parsimony (Kuhn, 1977; Popper, 1959). In addition, philosophers have discussed how models connect to the world in appropriate respects and to appropriate degrees; models with appropriate detail and complexity meet these purposes (Giere, 2004; Longino, 2002)</td>
</tr>
<tr>
<td>Complexity</td>
<td>Appeals to the appropriate complexity represented in good models, e.g., not too complex</td>
<td>10%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td>Models have sequence, e.g., has steps, it has a timeline, it has a cycle</td>
<td>29%</td>
<td>100%</td>
<td>Secondary. Ideas about sequence and steps invoke ideas about causation, mechanism, and explanation, all key elements of scientific models (Machamer et al., 2000). However, because students’ responses stopped short of referring to mechanisms, we assign it secondary rather than primary significance</td>
</tr>
<tr>
<td>Evidential criteria</td>
<td>Evidence</td>
<td>Explicitly refers to evidence supporting the model, using the word evidence</td>
<td>19%</td>
<td>93%</td>
</tr>
<tr>
<td></td>
<td>Other support</td>
<td>Refers to evidence supporting the model without using the word evidence, e.g., Data supports the model</td>
<td>8%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>Quality of support</td>
<td>Refers to the quality of evidence or reasons, e.g., the evidence should be true</td>
<td>6%</td>
<td>73%</td>
</tr>
<tr>
<td></td>
<td>Quantity of support</td>
<td>References the quantity of evidence or reasons, e.g., needs several pieces of evidence</td>
<td>2%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>Explicitly refers to reasons, e.g., gives reasoning.</td>
<td>9%</td>
<td>87%</td>
</tr>
<tr>
<td>Epistemic elements</td>
<td>Quantity of explanation, description, information</td>
<td>A model has a significant amount of explanation/descriptive information expressed in the model, e.g., it has a lot of explanation, explains lots of stuff</td>
<td>20%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>Appeals to an idea concerning how faithfully the model depicts the target, e.g., the model is correct</td>
<td>8%</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>Interest</td>
<td>Suggests that models should be interesting or creative</td>
<td>5%</td>
<td>53%</td>
</tr>
</tbody>
</table>
The explanation category suggests an understanding of explanation as a central goal of modeling. Because philosophers have identified explanation as an important goal of models (e.g., Giere, 1988; Kitcher, 1993), we classified this criterion as a primary epistemic criterion. In our coding scheme, we distinguished between two categories indicating that the purpose of models is to explain. The Explains-1 code (51% of all students) refers to the explicit use of the word explain (or morphologically derivative words such as explanation). The Explains-2 code (4% of all students) refers to other language used by the student that relates to explanation (e.g., a good model “shows how and why”). We differentiated between these two kinds of responses because our instructions at the beginning of the two-lesson sequence of activities explicitly mentioned that models provide explanations. Hence, although many students whose responses were coded as Explains-1 might have understood that models explain phenomena, others may have just been repeating what they had been told (but note that these students would have had to recall this information more than 24 hours after encountering it).

We characterized two types of goals (providing descriptions and modeling data) as secondary epistemic criteria. Forty-one percent of all students wrote that models provide descriptions. In the sciences, descriptive models play an important role, particularly in the applied sciences, and often serve as a foundation for the development of explanatory models. Descriptive models are also ubiquitous in science education materials (e.g., models identifying part of the cell, models identifying the location of planets), thus students may be very familiar with this goal of models. But description is less central to modern science than is explanation; as we noted above, in most fields scientists seek to develop explanatory frameworks. A smaller number of students suggested that models are data (5%)—as opposed to stating that models are supported by data. Data models are indeed a very important and specific kind of scientific model (Chinn & Brewer, 2001; Lehrer & Schauble, 2006; Staley, 2004), but they are distinct from explanatory models. Thus, we deemed ideas relating to models as data as a secondary level criterion for the quality of the explanatory models that students had been exposed to. It is unclear whether students providing this response had an understanding that data can be modeled or whether they were confused about the distinction between models and data supporting models.

Three responses suggested a vague or nonnormative understanding of models. Twenty-one percent of students indicated that models provide answers to questions, but the kinds of questions that students were referring to was unclear. We suspect that if these students had been prompted to elaborate or provide examples, the questions generated would be a mix of explanatory, descriptive, and narrow information-seeking questions. Responses in the information category (41% of all students) suggested that students believed models were providing static or isolated pieces of information. Of the students who provided

Table 1 (Continued)

<table>
<thead>
<tr>
<th>Criteria elements</th>
<th>Definition and example</th>
<th>Students (n = 324)</th>
<th>Classes (n = 15)</th>
<th>Interpretation of epistemic status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realism</td>
<td>Explicitly use the words real or realistic</td>
<td>2%</td>
<td>33%</td>
<td>Secondary. Philosophers have debated whether scientific models should be viewed in a realist, or at least a truthlike, sense (e.g., Niiniluoto, 2002). Students who give this criterion have, like many scientists, adopted a realist stance. However, exactly what students mean by “real” or “realistic” is unclear; they could mean that it is accurate, or that it has real entities, or simply that it is plausible. Hence, due to its relative lack of clarity, we have assigned this code to the secondary category</td>
</tr>
<tr>
<td>Quality</td>
<td>A model has a high quality explanation/description/information expressed in the model, e.g., a good model has a great explanation</td>
<td>19%</td>
<td>93%</td>
<td>Vague. Task instructions requested that students make a list of the most important characteristics of good (not bad) models. Students’ responses which imprecisely noted model goodness may have been simply repeating the task instructions</td>
</tr>
</tbody>
</table>
responses that fell within the information category, 80% proposed other primary and secondary criteria that suggested a deeper understanding of the goals of models (106 students). The third response suggesting a vague or nonnormative understanding of models was to view models as themselves being examples (8% of all students); this response seems compatible with a view of models as illustrations of explanations rather than as the actual proposed explanations. However, only 7% of the students who generated responses in the examples category (two students) failed to generate an additional type of model goal. Thus, most students who gave this response had a view of models that was not limited to models as examples.

In considering the entire corpus of criteria relating to model goals, students in three classes collectively proposed all six criteria, eight classes proposed five different criteria, three classes proposed four different criteria, and one class proposed three criteria. All classes had multiple students who advanced explanations and information as criteria. Thirteen of 15 classes had multiple students who advanced description as a goal. Thus, teachers who wish to draw on students’ ideas about the goals of models within a class discussion would likely find that a wide variety of model goals, including the most important primary and secondary goals, would be generated by one or more students in any given class.

In summary, a majority of students demonstrated some understanding of multiple model goals, and students in all classes collectively noted a wide range of goals. The diversity of student responses surprised us, given that previous research on students’ understanding of modeling has reported that most students see models only as direct replications of reality (Grosslight et al., 1991; Treagust et al., 2002). Judging from the number of students who suggested that models are examples, we suspect that some students see replica-type models as an important kind of model. However, this was a small percentage of students (only 8%). Our orientating activity may have steered students away from a misinterpretation of the word model as simply referring to the everyday usage of the word as a scale replica. The orientating activity presented students with models with a variety of different goals; the majority of these models were explanatory, but we also included descriptive models, replica models (which we do not count as scientific models), and even models of data.

Although students were introduced to the idea of scientific models with a definition that stated that models are explanations, we were still surprised by the number of students that referred to explanation in their criteria. One possible account for this finding is that students simply remembered and repeated what they were told about models at the beginning of the orientating activity on the previous day. However, our introductory definition also emphasized that models can be used to make predictions, and no students mentioned predictions or any related idea in their criteria. If students were simply repeating the introductory definition, they would have also mentioned ideas about predictions in their criteria. Thus, students either already knew something of the explanatory role of scientific models, or the idea of explanations fit so readily with their prior knowledge that they were immediately able to adopt this idea with minimal instruction.

Class level data show that all classes collectively put forward primary epistemic ideas. Thus, teachers who elect to lead class discussions about criteria can expect students to raise significant goal-related criteria, without having to suggest these criteria themselves.

Providing students with more opportunities to evaluate models with their classmates may encourage students to develop an appreciation for the epistemic value of explanatory models, as well as develop more nuanced criteria regarding purposes, such as “explanatory models are often more useful than descriptive models,” or a criterion that suggests that “the kind of model should match its purpose.” Criteria relating to the goals of models may be particularly easy for students to evaluate, as suggested by the high number of students who proposed criteria within this category. These questions should be explored in inquiry curricula.

Model Constituents

Scientific models take a variety of representational forms, including pictorial diagrams highlighting mechanisms, mathematical relationships, and even explanatory text. The constituent parts of these models—including color, titles, diagrams, and captions—aid in the important communicative function that models play when scientists present their work. Responses in the Model Constituents category specify features of models that help communicate what the models represent.

Eight criteria fell within this category. According to the three most common responses, students believed that models need pictures (60% of all students), words (30%), and diagrams (19%). Students also mentioned...
labels (13% of all students), arrows (12%), and titles (7%). Individual students mentioned an average of 1.5 model feature criteria. Twelve classes proposed six or more of these criteria; two classes proposed five; one class proposed four criteria within this category.

The connection between these constitutive parts of models and the epistemic aims of science is indirect. Without these parts, models would not exist; yet not all diagrams with arrows and titles are scientific models. For this reason, we think these criteria are best categorized as secondary epistemic criteria.

Taken as a whole, these criteria suggest that most students view models as something other than a literal or scale copy of an everyday object. A literal copy (e.g., a toy car being a model of a real car), does not require titles, arrows, and labels. But models that describe or explain complex phenomena would require these communicative tools to be properly represented.

As we have noted, in the orientating activity students evaluated whether the different representations of volcanoes and volcanic processes were or were not models; these representations varied in whether they had some of the constitutive features mentioned by the students. Students also viewed pairs of contrasting scientific models that included pictures, words, diagrams, and arrows. One specific example of a model students viewed and discussed was a picture representing the life cycle of a frog that did not include a title, label, or captions. Thus, we see a link between the orientating activity and students’ criteria, as students’ lists included constituents that were missing from this model.

Model constituent criteria are likely very accessible to most students. We think that these ideas might be a good starting point for introducing model critique activities to young students. As students become more adept at identifying and using these model features, they can progress to considering more central epistemic criteria. In addition, this category of responses reflects familiarity with the kinds of models students see in traditional instruction (e.g., pictorial or diagrammatic models in textbooks). As students are introduced to different kinds of models (e.g., mathematical models and explanatory models emphasizing unobserved mechanisms), new types of model features should emerge in criteria lists, and students may also realize that few, if any, features appear in all types of models. For instance, many students in our sample believed that good models require pictures; however, as students gain familiarity and skill at working with more abstract representations, they may come to believe that pictorial representations are not necessary, and in some cases can contain unnecessary and distracting details.

Communicative Elements

Criteria in the Communicative Elements category are closely related to those in the Model Constituents category. The Communicative Elements criteria reflect more general ideas about how models are typically designed in order to communicate ideas appropriately and clearly. The focus is not on the constituents of the models per se (which also play a communicative role) but in how the constituents are organized to make the models understandable to others. The three most common responses in this category concerned appropriate clarity (64% of students), appropriate details (40%), and the focus of model representations (37%). Responses classified as indicating a “focus of model representations” suggested that the model is designed specifically for the topic at hand. For example, if a model is supposed to depict photosynthesis, it should not instead depict transpiration.

Each of the seven criteria in this category was proposed at a relatively high frequency, with the average individual proposing 2.1 criteria involving communicative elements. Students in eight classes collectively proposed all seven criteria in this category. Four classes proposed six of the criteria, and three classes proposed five of the criteria, with complexity and audience being the only two criteria missing in these classes. Thus, teachers who wish to lead a discussion about criteria following individual generation of criteria can expect most or all of the communicative criteria in Table 1 to be mentioned by one or more students in each class.

Ideas about clarity, organization, focus, and audience are primarily communicative and apply to any kind of scientific model. Models cannot be evaluated for accuracy or fit with evidence if those who peruse them cannot make out what they mean; thus, models must meet criteria of successful communication before they can be evaluated for epistemic quality. Because a false or even fraudulent model can successfully meet communicative criteria, we judge that most communicative criteria are secondary rather than primary epistemic criteria. As social epistemologists such as Bishop and Trout (2005) have noted, the successful
spread of knowledge requires effective communication of ideas. A brilliant, empirically successful model that is presented in a way that other scientists cannot understand will not gain acceptance. Thus, communicative criteria are very important at the secondary level of epistemic practice.

Communicative criteria, especially criteria concerning clarity, organization, and focus are likely well established within school cultures. These criteria would also apply to tasks such as writing an essay in language arts or completing a project in history class. It is somewhat surprising that more students did not suggest these criteria. One possibility for why more students did propose these criteria is that many students considered them too obvious or not specific enough to the task of creating a list of criteria for good models.

Two criteria in this category, “models should have appropriate details” and “models should be appropriately complex,” seem to relate directly to ideas about parsimony. Parsimony is a primary epistemic criterion discussed in the philosophy of science literature on theory and model choice (e.g., Kuhn, 1977). A priori justifications of parsimony generally are centered on the idea that adherence to the parsimony criterion is one defensible way to work around the problem that models are underdetermined by supporting data. Naturalized approaches to epistemology point to a number of important episodes in science in which scientists explicitly adhered to this criterion (e.g., development of evolutionary explanations; Sober, 1981). We note, however, that although students proposed criteria for the value of appropriate details and complexity, they did not explicitly mention the value of idealizations. Actual scientific models are idealizations that capture only those aspects of a phenomenon that are of interest to the modeler. Thus, they do not contain pictorial details that are irrelevant to the model purposes (e.g., a model of photosynthesis processes in a microbiology journal does not include a picture of the flowering parts of the plants, which are irrelevant to photosynthesis). When students mentioned appropriate details as criteria, it is not entirely clear whether they were advocating leaving out details that are hard for the audience to understand or details that are unimportant for the purposes of the model.

Students also frequently mentioned the importance of sequence in models (29% of all students). Sequence evokes ideas related to causation. Developmental research has shown that even very young children can invoke causal mechanism when describing how objects work (e.g., Wellman & Gelman, 1992). Temporal and visual contiguity, as would be found in a pictorial or written sequence, have been found to encourage reasoners to invoke causal explanations (White, 1988). This provides more support that many, if not most, students recognized the explanatory role of scientific models. We have categorized the sequence criterion as a secondary epistemic criterion because there is only an indirect link to causation and explanation. As students gain more experience with causal and explanatory models, we expect that students will begin to articulate the conditions for good causal mechanisms in criteria.

A number of the criteria within the Communicative Elements category appear to relate directly to the contrasting models that students viewed during the orienting activity. For example, students compared a complicated food web showing the relationships among 13 species with a simple food web that only showed three species. These contrasting models might have spurred ideas about appropriate levels of simplicity or complexity. Similarly, students compared a water cycle model that presented only relevant, well-defined structures and mechanisms, with one that included additional, unnecessary and cute pictorial details serving only an aesthetic function (i.e., cartoon images of rain, lakes, clouds, and a landscape which included houses). These models may have spurred students’ ideas about appropriate detail. These models also all had a strong sequential component, which could have encouraged students to articulate sequence as a criterion.

Overall, the high frequency and relative sophistication of the proposed communicative criteria suggest that, collectively, students begin with ideas that provide a foundation for developing more nuanced criteria. For example, students’ ideas about model complexity and detail can be the basis for discussions about the appropriate degree of parsimony. Similarly, instruction could aim to move students’ vague ideas about sequence to more explicit criteria describing what a good mechanism entails. Thus, students’ rich preinstructional ideas provide resources for developing more nuanced understandings. Similarly, there are a number of well-established “school” criteria that students proposed, such as clarity. It may be appropriate to discuss these criteria when introducing model criteria, but because most students are familiar with them, teachers may want to emphasize these criteria less than criteria that are less familiar to students. Explicit meaning-making discussion should likely be centered on primary and secondary criteria that are less established in students’ repertoires.
Evidential Criteria

The category evidential criteria comprises responses that referenced evidence and reasoning. Evidence is central to the enterprise of model-based science (Giere, 1988; Longino, 1990, 2002; Niiniluoto, 2002). The epistemic justification for a model is ideally, and most commonly, based on the quantity and quality of evidence supporting the model. Further, a model can be questioned if there is evidence that contradicts the model. Thus, responses within this category reflect awareness of model goodness criteria that are central to the epistemic practices of science.

Overall, 24% of students directly indicated that good models are supported by evidence. Most of these students (19% of all students) used the word evidence; 8% of all students used other language that suggested understanding of evidential support. Three percent of students wrote one criterion in each of these two categories. Nine percent of students noted that models should include reasons or reasoning. A few students mentioned the importance of the quality and quantity of evidence or reasons (6% and 2% of students, respectively). The average individual proposed 0.4 criterion related to evidence.

All but one class included at least one student who proposed a criterion that used the word evidence. The majority of classes had at least one student who used other language to suggest evidence, noted ideas about the quality of evidence, or mentioned reasoning. In 43% of classes students proposed criteria concerning the amount of evidence. Eighty-seven percent of classes had at least three students who proposed one of the criteria in the evidential criteria category.

The relatively high use of the words evidence and reasoning suggest that many of these middle school students had some awareness of the role of evidence in epistemic practices. This is a particularly interesting result because our orientating activity did not make any direct reference to the role of evidence in evaluating models and because research suggests that many students struggle to explicitly differentiate between evidence and theories or models (Zimmerman, 2007). Had we inserted some rudimentary evidence or ideas about evidence into our orientating activity, perhaps an even larger number of students would have specified a role for evidence in evaluating models.

As discussed above, previous studies have provided conflicting results on students' conceptions of modeling revolved around the role of evidence in modeling (Grosslight et al., 1991; Treagust et al., 2002). The results of this study are congruent with Treagust et al.’s (2002) finding that students have some understanding of the role of evidence in model evaluation. Treagust et al. (2002) found that most students agreed with statements suggesting that models are to be revised in light of evidence. Our measure was not sensitive to whether students’ simply understood that evidence was somehow involved in modeling practice, that evidence is used to choose between models, or that models can actually be revised in light of evidence. Our study shows that many students spontaneously generate criteria related to evidence, without prompting.

Students’ criteria concerning reasons and reasoning were vague. Students likely have many ideas about what constitutes a reason. A reason could tout the logical coherence of a model, note a model incoherence, or it could specify how evidence supports a model. A reason could also be nonepistemic grounds for accepting or rejecting a model. For example, Hogan and Maglienti (2001) found that students appealed to their own prior beliefs when evaluating conclusions, rather than drawing on provided evidence. Within the context of this task, we think that responses concerning reasons and reasoning suggest that some students have an emerging understanding that scientific models are not static artifacts, but are developed and used in concert with evaluation by scientists (or students within an inquiry environment).

As discussed previously, Kuhn (1977) emphasized the importance of criteria for the accuracy and explanatory scope of a model. Model quality is higher when there is a good fit between models and evidence. The presence of ideas about evidence in students’ lists of criteria indicates that many students are aware of normative criteria of accuracy with respect to evidence. A few of the middle-school students even displayed an even more nuanced view of evidence by appealing to the quality and quantity of evidence. There is no reason, of course, to think that students’ understanding of accuracy is as nuanced as philosophers or scientists. For example, there was no indication that they understood evidential fit in terms of the closeness of quantitative fit with the predictions of a mathematical model. Nor did they specifically identify covering a diversity of evidence types as an important criterion (e.g., the success of evolutionary theory in explaining phenomena in many different domains, from molecular biology to paleontology). Nonetheless,
there is evidence that some students have developed some significant precursors of these more sophisticated ideas.

Epistemic Elements

The Epistemic Elements category consists of five generic responses relating to model quality. Twenty percent of students wrote about the quantity of explanations, descriptions, and information presented in models (e.g., a good model has lots of explanation), 8% of students proposed criteria for accuracy, 5% proposed that good models are interesting or creative, 2% proposed criteria referencing ideas about realism, and 19% of the students referred simply to model quality in a general way. On average students proposed 0.5 criteria in this category. Most classes had at least one student who proposed criteria concerning quantity (93% of classes), accuracy (80% of classes), and interest/creativity (53% of classes). Only a third of the classes had students who proposed a criterion relating to model realism.

Students who mentioned model quality may have simply been echoing the task instructions to make a list of the most important characteristics of good (not bad) models. For this reason, we have classified criteria concerning general model quality as being too vague to be assigned to either primary or secondary epistemic status.

Students who proposed the four other criteria within this category appear to have demonstrated an emerging grasp of several normative epistemic criteria. For instance, students who proposed criteria relating to the quantity (e.g., a good model has lots of explanation) may have held the normative conception that good models can explain a large array of phenomena and observations (Kuhn, 1977). We viewed this criterion as a primary epistemic criterion, due to the centrality of far-reaching explanations in science. Another interpretation is that students believed that a model is better to the extent that students (or scientists) have invested a greater quantity of work or effort into constructing or working on a model. This interpretation is more consistent with existing school norms, but because no students used additional language to support this interpretation, it is not well supported by overall pattern of results.

Students in this study, much like philosophers of science, wrote that models should be accurate. It is possible that these students were thinking of a more literal mapping between the model and the phenomena (Grosslight et al., 1991; Treagust et al., 2002), but it is also possible that students were referring to evidential fit of the sort discussed in the previous section. It is possible that students were simply giving voice to well-established school norm that students generate a factually correct response that matches the teacher’s expected answer, such as when answering an algebra question, responding to a history teachers’ question, or completing science work in a traditional classroom (Duschl, 1990). Despite this possibility, we classified accuracy as a primary epistemic criterion due to its close association with fit with the evidence, as well as its implied association with truth, which many philosophers have taken to be a condition of knowledge (Alston, 2005). Moreover, we think it is more likely that if students were focused on teacher-provided or sanctioned ideas, rather than empirically supported ideas, they would be more likely to write that the model was “right” or that it was the “right answer.” This is a case where teachers (and researchers) should collect more detailed information on students’ conceptions of accuracy, as well as aim to move students from the well-entrenched school conception to a scientific conception.

The students who proposed that good models are interesting or creative expressed an idea that philosophers have discussed as very important to the functioning of science. A model that is interesting to a student is probably presenting new, unexpected information to them. It appears to us that this type of criterion captures a sense of novelty that is central to scientific progress, and it also reflects the importance of uncovering not just any truth but significant truths that are of interest to the scientific community (Haack, 2003; Kitcher, 2001). We thus count this as a central epistemic criterion.

Finally, students who wrote that good models are real or realistic, may be demonstrating a more simplistic notion of models being literal replications of phenomena, or they may be expressing an idea that models should be accurate in reflecting reality in some way. Philosophers (e.g., Cartwright, 1983; Giere, 2004; Kuhn, 1977; van Fraassen, 1980) have debated scientific realism, that is, the question of whether the theoretical entities and processes of science refer in some way to counterparts in the world. Many scientists have adopted critical realist stances in which models are viewed as truth like or resembling the world in important respects (Giere, 2006; Niiniluoto, 2002). Thus, students’ ideas about the realism of models may not
be that different from many scientists. However, given that students did not express exactly what they meant by this criterion, we count this as a secondary epistemic criterion.

While completing the orientating activity, students evaluated two pairs of models that potentially contributed to the incubation of ideas related to the criteria within this category. For example, they evaluated two models representing diffusion, one depicting Brownian motion and the second depicting a non-normative model in which molecules reproduced and multiplied as if they were living things. From this pair, students may have recognized that it was unlikely that both models could be accurate, although most were probably unaware of which model was closer to the scientifically accepted model. Similarly, the two food web models, described above, may also have contributed to the development of these criteria; students may have decided that the more complex model explained a larger part of the ecosystem, and thus was more accurate or perhaps “real.”

**Coordinating Responses Across Categories**

A question raised by our data is the extent to which the most sophisticated responses were made by the same individuals. Was it the case, for example, that a small number of sophisticated students proposed the majority of the primary criteria? Or were the more normative responses spread across a broader range of students?

Of the 31 criteria in our coding scheme, we categorized ten criteria as suggesting an understanding of primary epistemic aspects of science, 17 criteria as suggesting an understanding of secondary epistemic aspects of modeling, and four which were vague or suggested a non-normative understanding of modeling. According to our scheme, a student with the highest level of epistemic understanding would propose primary epistemic criteria suggesting that scientific models (a) provide explanations (explains-1 or explains-2), (b) require evidence (generated at least one code in the evidential criteria category), (c) are interesting, (d) represent appropriate detail or complexity, (e) have significant scope (quantity code), and (f) are accurate. No student proposed criteria within all six of these categories, and only 2% of students proposed criteria that fell into five categories. Eight percent of students proposed criteria in four categories and 18% proposed criteria which fell into three of these main categories. Thus, all students in our sample had room for significant epistemic growth at the individual level. We did find that 16% of all students proposed both explanation and evidence, perhaps the most important of the primary epistemic criteria.

It is also of interest to examine the distribution of vague criteria. A vague set of personal criteria would not allow students to effectively evaluate peer or teacher-provided models. We found that 34% of student had no vague criteria on their list. Twenty percent of students had over a quarter of their criteria categorized as vague. Only five students (1.5%) had half or more of their criteria coded as vague or non-normative. Thus, most students had some understanding of primary and secondary criteria which were not vague.

Finally, we examined the distribution of responses across the five categories (goals of models, model constituents, communicative elements, evidential criteria, and other epistemic criteria). Overall, seven percent of students proposed criteria that fell into all five categories of criteria. Thirty-four percent proposed criteria within four categories and 43% proposed criteria in three categories. Thus, most students suggested criteria that spanned a fairly broad range of epistemic categories.

**Conclusions**

Our goal in this analysis was to explore middle school students’ ideas about criteria for judging the quality of scientific models. We extended previous research by both focusing on students’ own ideas as they developed lists of epistemic criteria and by interpreting these results at a collective as well as an individual level. Students demonstrated familiarity with a wider range of modeling ideas than has been documented by previous studies. Although many students proposed established, school criteria, these were seldom the dominant type of criteria on students’ lists. At a class level, most classes had multiple students who proposed primary, normative criteria.

**Student Generated Epistemic Criteria**

Students’ criteria provided insights into their understanding of the goals of models, constituent parts, communicative features, the role of evidence in modeling, as well as other epistemic features.
A number of epistemic criteria correspond to those used by practicing scientists. Of five prominent scientific criteria identified by Kuhn (1977), students generated ideas which resembled three:

- **Accuracy**: Eight percent of individuals and 80% of classes specifically noted accuracy as a criterion, while 24% of individuals and all 15 classes proposed criteria relating to evidence, which also appears to be related to the accuracy of models.
- **Explanatory scope**: Twenty percent of individuals and 100% of classes noted the importance of quantity of explanations/descriptions/information.
- **Parsimony**: Over 40% of individuals and 100% of classes proposed that models should have appropriate details or complexity.

Students did not generate any criteria related to Kuhn’s fourth and fifth criteria: internal or external consistency, or fruitfulness. Students also proposed seven high-level communicative criteria (e.g., *models are clear, appropriately detailed, etc.*), which philosophers have argued play a central role in the uptake of scientific ideas (Bishop and Trout, 2005). Considering that the students had limited experience with scientific models and that they had not engaged in discipline-specific model-based inquiry, we conclude that the students displayed a relatively sophisticated understanding of communicative criteria for good models.

There are, of course, a number of criteria identified by philosophers of science that students did not generate and that could be introduced to students over time. Some of these include criteria relating to observational nesting (i.e., a more successful model explains data that competing models explain and successfully explains additional data, as well), the historical track record of empirical success, and the smoothness with which adjustments can be made in the face of explanatory failure (i.e., a model is preferred if its explanatory failures can be corrected with minimal modifications) (Newton-Smith, 1981).

Our results differ from some of the previous research on students’ naïve understanding of both the nature of science and modeling practices. For example, Grosslight et al. (1991) found that none of the seventh graders in their study elaborated on the role of evidence in constructing models; in contrast, many participants in the present study spontaneously noted that evidence was important. Both Grosslight et al. (1991) and Treagust et al. (2002) concluded that students primarily see models as direct replicas of phenomena rather than abstract or idealized explanations. Students in the present study produced criteria that suggested that many understood a model’s explanatory function, by explicitly noting that models are explanations and that models often include representational devices, such as arrows, titles, and labels to explicate models. If students saw models only as scale, literal representations, they would have not seen these devices as necessary.

A conservative interpretation of our results is that students were merely able to identify appropriate vocabulary that could be used to generate model quality criteria. If this were the case, students’ understanding of modeling and epistemic criteria could be interpreted as superficial and unsophisticated (Grosslight et al., 1991). However, this interpretation is not supported by the overall pattern of results. As discussed previously, only one vocabulary word (*explanations*) appeared both in the orientating activity and in students’ criteria. Another important vocabulary word (*predictions*) appeared in the orientating activity (in the same paragraph with the word *explanation*), but no students generated criteria related to predictions. In addition, student used terms such as evidence, accuracy, complexity, and sequence, none of which appeared in materials that they had viewed. Therefore, a better-supported interpretation of the results is that students exhibited an understanding of a broad variety of criteria for model goodness, albeit with a lesser emphasis on criteria related to evidential support than on other criteria.

One possible reason for the better understanding of students in this study is that we provided them with more authentic examples of scientific models to consider as they generated their criteria for model quality. Grosslight et al. (1991) presented students with a toy airplane, a map, a static picture of a house, and a textbook diagram of the water cycle early in their interview protocol. Treagust et al. (2002) did not provide participants with any examples of scientific models before they completed a questionnaire. Participants in these two studies may have been confused about the kind of models they were evaluating; it seems likely that a toy model would be the first kind of model that comes to mind for a seventh grader, and even for many adults. Hence, participants may have assumed that toy models were the sort of thing that the researchers were asking...
about. In contrast, our task may have allowed students to grasp what sort of thing is meant by the term "scientific model," rather than requiring them to guess. Although we think that a better understanding of the task improved students’ responses, it is unlikely that our participating students had had many substantive opportunities to engage with or evaluate scientific models. Thus, rather than drawing on any explicitly learned criteria for models, students were developing and crystallizing their criteria as they reflected on the models they were given to analyze.

Students’ responses appear to have been associated with the specific models they evaluated in the orientating activity. We think that this suggests one of two possibilities. One possibility is that seventh-grade students have existing ideas or resources for evaluating epistemic artifacts, honed during non-epistemic evaluative activities. Specifically, students may have general “evaluative resources” developed through everyday activities such as judging favorite music, fiction, or choosing a best or favorite superstar soccer player. Students have few opportunities to engage with these ideas during epistemic practices such as choosing between alternate scientific models; the model evaluation task may have prepared students to restructure their existing evaluative resources around the epistemic aim of evaluating models (Louca et al., 2004). Recent research has examined how different contexts, activities, and instructions may lead students to approach activities in very different ways. For example, Engle, Nguyen, and Mendelson (in press) concluded that transfer (of learning) was facilitated when instructors adopted an “expansive frame,” characterized by encouraging students to think beyond the specific learning task. Similarly, Scherr and Hammer (2009) interpreted particular student behaviors and activities as being associated with multiple epistemological frames (e.g., an opportunity for sense-making vs. seeing an assignment only as busywork). The orientating activity may have appropriately reframed the criteria generation task for students such that it highlighted the types of knowledge and norms which were most appropriate for the task.

The second possibility is that students constructed entirely new knowledge about models and model evaluation through the orientating task and from interacting with their peers. However, this conclusion is not well supported. We have only been able to speculate about possible relationships between the models students viewed and the criteria they generated; we saw no explicit links in their written criteria. In addition, if students’ criteria were primarily derived from the model evaluation task, fewer students would have proposed criteria relating to evidence and more students would have generated criteria related to explanations. One limitation of the current study is that our data are not robust enough to link students’ individual epistemic criteria directly to their engagement with the models that they viewed. The orientating model-evaluation activity, despite being relatively short, provided a context for thinking about models. We think that students might have become more attuned to evidentiary considerations had the task included more models that students could recognize as being consistent or inconsistent with familiar evidence. An interesting topic for future research is to investigate how variations in the exemplar models to which students are exposed affect the criteria that they generate. In addition, future research should examine criteria generated by students in different age bands, as well as criteria that are manifested during processes of learning and inquiry (cf. Sandoval, 2005; Greene & Azevedo, 2010; Greene, Muis, & Pieschl, 2010).

Criteria and Instruction

Given the rich array of criteria proposed by students, the results of this study provide important insights into the kinds of instruction that may improve students’ model-based reasoning and understanding of the nature of science. The task of generating epistemic criteria would likely provide teachers with information on their own students’ level of understanding with which they can make instructional decisions.

If students can generate reasonable sets of criteria on their own, then it would be possible for students’ criteria to become the basis for classes to construct community norms to use epistemic and other criteria for model goodness. In our view, the students in the current study did indeed generate a reasonable set of collective criteria. Although few individuals could generate most primary epistemic criteria on their own, nearly all primary epistemic criteria were evident in each of the 15 classes in the present study. Therefore, the goal of instruction should perhaps focus on facilitating the sharing and vetting of these epistemic ideas. Social mechanisms may foster wide adoption of high-level epistemic criteria in the classroom. Anderson et al. (2001) found that when argumentation strategies are used in discussion, they tend to spread and occur with increasing frequency in classrooms with open-participation formats (as opposed to teacher-led discussion.
formats). The use and understanding of high-level criteria may spread and increase in a similar fashion. Student proponents of criteria could be encouraged to explain their rationale during discussion about criteria. Such discussions could help other students understand and appreciate the importance of specific criteria; further, students will be engaging in precisely the kind of dialogic process that scientists partake in when proposing and vetting criteria. This could develop nuanced understandings of normative epistemic standards, while also developing knowledge of broader community processes for establishing and vetting standards.

Centering inquiry on students’ own epistemic criteria may support the development of powerful classroom learning communities. Engle and Conant (2002) described a science classroom in which students were held accountable to disciplinary norms (e.g., sharing ideas and using evidence); observations suggested adherence to these norms led to greater authentic, disciplinary engagement. In general, the norms were established through teacher introduction, emphasis, highlighting, and encouragement. Having students generate their own epistemic standards, as we have done, may encourage the rapid transfer of this responsibility from teacher to students, by giving students a sense of ownership over criteria and assuring that at least some students understand the criteria.

The processes of sharing and using criteria could be scaffolded in a number of ways. First classes of students can work together to develop a class list of criteria. This class list could serve as the foundation for emerging epistemic norms in the classroom. Students can revisit and revise this list as their ideas become more normative. This list would serve as a public record of the students’ collective, emerging understanding of criteria. Second, students and teachers can develop rubrics based upon criteria (drawn from the best or most common student ideas, from a class list of criteria, or even from the list described in this study). Students can then use the rubrics to evaluate both their own and provided models; this may lead to deeper content knowledge. Finally, having students rank criteria may help students identify key differences in epistemic levels (e.g., the differences between what we have dubbed as primary and secondary criteria). Future research should explore the efficacy of these approaches.

We think that epistemic criteria may serve as a tool with which to develop students’ understanding of nature of science (NOS). Most studies of students’ understanding of NOS suggest that students do not see theories or models as tentative and explanatory, nor do they grasp the dialogic processes of coordinating theories with evidence (Lederman et al., 2002). In this study, however, we have found that some students understand that fit with evidence is a criterion for models and that good models are explanatory. These incipient understandings of criteria for good models could be the basis for discussing the nature of science and encouraging a more sophisticated NOS understanding. Students’ initial criteria for good evidence and for good arguments could serve a similar role in discussions about the role of evidence and arguments in science.

As with most new instructional practices in science, engaging students in modeling in a sophisticated way is challenging because it is inconsistent with many well-established school, classroom and instructional norms (cf. Chinn & Malhotra, 2002; National Research Council, 2007). One of the attractions of making students’ own criteria the starting point for model-based instruction is that it provides students with a high degree of authority and ownership over the classroom norms (Blumenfeld, Kempler, & Krajcik, 2006; Engle & Conant, 2002). Most inquiry programs attempt to foster student authority during the process of deciding on questions or developing intellectual products (e.g., models and experiments). Given that classroom norms shape and scaffold students’ choices when generating inquiry questions or developing products, epistemic criteria seem to be an advantageous starting point. This research suggests that students have the resources to develop a reasonable set of criteria. Future research should explore to what extent students use their own criteria and at what point students recognize the limitation of their criteria and amend them.

Summary

Science students are often thought to have unsophisticated knowledge of the epistemic practices of science. We found, to the contrary, that students seem to have a wide range of ideas about one important element of this practice—the epistemic criteria for good models. Many of the criteria proposed by students are similar to the criteria used by scientists (as identified by philosophers of science). Students collectively generated 31 distinct criteria for evaluating scientific models. Primary epistemic criteria included criteria
related to the explanatory function of models, the role of evidence, appropriate details, and accuracy. Fifteen criteria were related to the communicative or constituent features of models; these suggested that students saw models as more than just direct copies or scale models. Students generated several different evidentiary criteria, although only about a quarter of individuals generated these criteria. Given the complexity of student responses, we think epistemic criteria are a potentially powerful starting point for instruction aimed towards improving student reasoning and understanding of the nature of science. The distribution of student responses within classes suggest that instruction centered on peer sharing and vetting of ideas related to criteria would be an appropriate instructional strategy.

This material is based upon work supported by the National Science Foundation under Grant No. 0529582. 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 National Science Foundation. We thank Richard Duschl for his many contributions to the PRACCIS project and Luke Buckland for his very helpful comments on earlier drafts of this paper.

References


Journal of Research in Science Teaching


