Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations

MARK WINDSCHITL, JESSICA THOMPSON, MELISSA BRAATEN
Curriculum and Instruction, University of Washington, 115 Miller Hall, Box 353600, Seattle, WA 98195, USA

Received 19 July 2007; revised 21 November 2007; accepted 29 November 2007

DOI 10.1002/sce.20259
Published online in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: One hundred years after its conception, the scientific method continues to reinforce a kind of cultural lore about what it means to participate in inquiry. As commonly implemented in venues ranging from middle school classrooms to undergraduate laboratories, it emphasizes the testing of predictions rather than ideas, focuses learners on material activity at the expense of deep subject matter understanding, and lacks epistemic framing relevant to the discipline. While critiques of the scientific method are not new, its cumulative effects on learners' conceptions of science have not been clearly articulated. We discuss these effects using findings from a series of five studies with degree-holding graduates of our educational system who were preparing to enter the teaching profession and apprentice their own young learners into unproblematic images of how science is done. We then offer an alternative vision for investigative science—model-based inquiry (MBI)—as a system of activity and discourse that engages learners more deeply with content and embodies five epistemic characteristics of scientific knowledge: that ideas represented in the form of models are testable, revisable, explanatory, conjectural, and generative. We represent MBI as an interconnected set of classroom conversations and provide examples of its implementation and its limitations. © 2008 Wiley Periodicals, Inc. *Sci Ed* 1–27, 2008

Correspondence to: Mark Windschitl; e-mail: mwind@u.washington.edu.

INTRODUCTION

Activity without understanding seems to be a regular feature of classroom life for science students in American schools (Banilower, Smith, Weiss & Pasley, 2006; Roth & Garnier, 2007). Reversing this state of affairs will be among the major projects facing science education over the next 10 years; specifically this will involve the reenvisioning of implementable frameworks for inquiry in K-16 settings that engage learners deeply with content and with the epistemic practices of authentic science. In this essay, we propose that forms of inquiry based on the generation, testing, and revision of scientific models—i.e., model-based inquiry (MBI)—can support this kind of learning at all levels of education.

Any plans for testing new frameworks for school science investigations would be poorly conceived, however, without taking into account the influence of the current paradigm of preference for educators—the scientific method (TSM)—and its role in allowing distorted images of science to be passed down through schooling practices. We provide an illustration of this phenomenon from studies with beginning teachers who had earned undergraduate degrees in science and were preparing to communicate how they understood science to the next generation of learners. For those of us who assume that exposure to college-level science and to scientists themselves as instructors would gradually erode one's belief in the simplified methods and narrow epistemology associated with TSM, we provide evidence to the contrary—that a faith in unproblematic procedural processes is not only alive and well but dominates the conceptual frameworks that both novice and experienced teachers use to design investigative science experiences for students.

If it is difficult to escape this kind of history, it is impossible to ignore culture. Reference to a universal scientific method is common in discourse at all levels of science education. One needs only to open popular science texts, to examine students' laboratory notebooks, or to listen to science fair presentations at your local middle school to recognize that TSM remains a durable icon that actively shapes how teachers and learners think about scientific practice (see Abell & Smith, 1994; Bencze & Bowen, 2001; Bencze & Hodson, 1999; Brickhouse, 1990; Lederman, 1992; Palmquist & Finley, 1997; Simmons et al., 1999; Solomon, Duveen, & Scott, 1994). Most teachers and many of their students can recite from memory the steps of this process. With only minor variations: observe, develop a question, develop a hypothesis, conduct an experiment, analyze data, state conclusions, generate new questions. We assert, however, that TSM is not scientific at all when considered from an epistemic perspective, and that it subverts young learners' understandings of both the practices and the content of the discipline.

Many educators would take these as outrageous claims. How can a framework that has guided teachers and students to think in systematic ways about observation, questioning, and experimentation be anything but scientific? How could it possibly impede young learners' abilities to reason about the natural world? In response we argue from two positions. One is rooted in the disciplinary epistemology of science, and how it defines the kinds of human activity that generate and validate scientific knowledge. The other position is empirical, based on a series of five studies with different cohorts of aspiring teachers who had just finished their undergraduate degrees and were preparing to enter their own classrooms. From this combined record of more than 60 in-depth cases of preservice teachers, we identify patterns of incongruity and outright conflict between their conceptions of inquiry—constrained in large part by a history of involvement with TSM—and the practices of authentic science.

We recognize that critiques of TSM are not new. Rudolph (2005), for example, chronicled the spread of TSM as a movement that began with a simple five-step heuristic for "logical thinking" developed by John Dewey (1910). During the 20th century, this idea swept through the educational community despite regular admonitions from notable scientists (and Dewey himself) that there was no such thing as "a" scientific method. From a methodological perspective, Bauer (1992) picked apart the notion of a universal method, citing the varied ways in which members of different subdisciplines in science pose questions, acquire data, deal with theory, and argue with evidence. From a curricular perspective, Hodson (1996) traced the changing nature of science inquiry in schools from the 1960s to the present, arguing that movements such as discovery learning, process approaches, and particular forms of constructivist pedagogy have all misrepresented the nature of investigative science.

We do not aim to recapitulate these critiques here, rather we argue specifically for an alternative to the scientific method, which we refer to as model-based inquiry—a framework for school science that can be articulated in epistemically grounded terms. To do this, we must first identify core knowledge producing activities that scientists engage in, then consider current classroom practice in light of these. However, before offering an alternative to TSM we feel it is crucial to present evidence of the cumulative impact on science learners' exposure to limited views of investigative science. This research not only helps explain how TSM endures in our educational system but identifies specific challenges we will face in fostering more accurate conceptions of science for beginning teachers. As our goal, we do not seek to replace one formalism with another, but to advocate for comprehensible new orientations to inquiry science that can bridge the worlds of science and K-16 education in an intellectually honest way.

WHAT IS AUTHENTIC PRACTICE?

Five Features of Scientific Knowledge

Scientists engage in a wide range of activities. They watch other members of their profession perform demonstrations of new equipment and techniques, they build laboratory skills over time (e.g., safety practices, using equipment, learning specific procedures), they try to replicate other scientists' studies, they invent new technologies, conduct thought experiments, do library research, and use knowledge to solve practical problems (Bauer, 1992; Kosso, 1992; Pickering, 1995; Rouse, 1987). All of these activities are valuable for school science learning as well. But there are particular practices that are integral to the core work of science—this core being organized around the *development of evidence-based explanations of the way the natural world works* (Giere, 1991; Longino, 1990). Roughly speaking, this involves the creative process of developing hypotheses from theories or models and testing these against evidence derived from observation and experiment. Our conception of "core work" is not intended to represent specific activities that *all* scientists engage in, rather it is conceptualized around an epistemology of scientific knowledge held by the scientific community—a set of commitments about the nature and grounds of knowledge that help define "what counts" as a scientific way of generating and validating new ideas. Using this premise to reconceptualize school science inquiry, we advocate for coordinating the language and activities in classrooms around five epistemic features of scientific knowledge (Smith, Maclin, Houghton, & Hennessey, 2000; Windschitl, Thompson, & Braaten, 2007)—that it is

- testable,
- revisable,
- explanatory,
- conjectural, and
- generative.

By *testable*, we mean that scientific knowledge, in the form of models or theories, is advanced by proposing new hypotheses that express possible relationships between events, processes, or properties within these models or theories, and by using various domain-specific methods for gathering data aimed at evaluating these hypotheses. By *revisable*, we mean that scientific ideas can change in response to new evidence or because a phenomenon is conceptualized in an entirely different way (e.g., kinetic vs. caloric models of heat transfer). By *explanatory*, we mean that the goal of science is to provide causal accounts of events and processes, as opposed to accumulating descriptive detail about phenomena or merely seeking patterns. By *conjectural*, we mean that causal accounts often involve theoretical or unobservable processes that can only be inferred from empirical observation (data) and that scientific argument aims to persuade others that explanations based on these inferences account most adequately for the observations. By *generative*, we mean that scientific knowledge, in the forms of models and theories, are the prime catalysts for new predictions, insights about phenomena, and hypotheses for testing; they are not simply "end-products" of inquiry (Hempel, 1966; Knorr-Cetina, 1999; Kuhn, 1970; Latour, 1999; Longino, 1990; Ochs, Jacoby, & Gonzales, 1994). One can find all five of these characteristics of disciplinary knowledge interwoven through the activity and rhetoric of the great scientific projects of the past 300 years, including Newton's advances in optics, Mendel's experiments with inheritance, the development of plate tectonic theory by Wegener and others, and Einstein's work on relativity.

Modeling in Science and in School Science

Authentic forms of inquiry for school science can be grounded in these five ideas and, in particular, in reasoning with and about models. This is because epistemic issues such as testability, conjecture, explanation, principled revision, and generativity are embodied most clearly and commonly in scientists' (and students') work with models as representations of ideas about the natural world.

Contemporary studies of scientific work have demonstrated that experimentation and the broader enterprise of inquiry are becoming routinely situated in model building, testing, and revising (Darden, 1991; Duschl & Grandy, 2008; Giere, 1988; Kitcher, 1993; Longino, 1990; Nersessian, 2005). Scientists as a matter of practice express ideas in the forms of inscriptions (i.e., textual, pictorial, or graphic expressions such as notations, drawings, diagrams, charts, or maps), analogies, physical constructions, or computer simulations to describe and understand natural processes that range from atomic bonding to the life cycles of stars (Latour, 1990). Models can represent theoretical structures (as in abstract/conceptual) such as energy pyramids in ecosystems, or those simply inaccessible to direct observation, such as the interior of the earth. While some theorists have distinguished between models as "part of a natural process used to construct an explanation of a phenomenon" and models as "representational tools for communicating a conceptual referent" (Romberg, Carpenter, & Kwako, 2005), we share a synthesis view that models are representations constructed as conventions within a community to support disciplinary activity (Lehrer & Schauble, 2003). Regardless of how models are conceptualized, they generally emerge from some phenomenological context (event, question, problem), they involve identifying key features or attributes of the phenomenon, and they specify how they are related (Romberg et al., 2005). The general aim of modeling is to test an idea—represented as a system of related processes, events, or structures—against observations in the real world and to assess the adequacy of the representation (i.e., model) against standards of evidence. Successful instructional frameworks for modeling typically guide students through a number of processes: engaging with a question or problem (often through material involvement with a natural phenomenon); developing a tentative model or hypothesis about causal or otherwise associative relationships in the phenomenon; making systematic observations to test these hypotheses; creating models of the phenomena that would account for the observations; evaluating this model against standards of usefulness, predictive power, or explanatory adequacy; and finally, revising the model and applying it in new situations (see Hestenes, 1992; Lehrer & Schauble, 2006; Lesh, Hoover, Hole, Kelly, & Post, 2000; Metcalf, Krajcik, & Soloway, 2000; Schwarz & White, 2005; Stewart, Hafner, Johnson, & Finkel, 1992; Stewart, Passmore, Cartier, Rudolph, & Donovan, 2005).

Certain forms of MBI encompass all five dimensions of the epistemic nature of knowledge (noted in italics below). In these inquiries, models are treated as subsets of larger, more comprehensive systems of explanation (i.e., theories) that provide crucial frames of reference to help *generate* hypotheses for *testing,* act as referents in interpreting observations, and are themselves targets of *revision*. Arguments for the support of *conjectural* models (as opposed to purely descriptive, empirical models) involve using observations (e.g., a balloon affixed to the top of a flask will begin to inflate when the flask is heated) to support *explanations* involving unobservable entities or processes (in this case, that heat causes molecules of air to move more rapidly, producing a pressure inside the flask that is greater than that outside). Conjectural models are representations of scientific phenomena (inscriptions typically) that coordinate observable features of that phenomenon with hypothesized explanatory events, properties, or structures that are not directly observable because of their inaccessibility (the mantle of the earth or molecular movement) or conceptual nature (forces, metabolic rates in organisms).

In sum, although different domains in science have their own fundamental questions, methods, and standards for "what counts" as evidence, they are all engaged in the same knowledge-building pursuit—the development of coherent and comprehensive explanations through the testing of models (Hempel, 1966; Knorr-Cetina, 1999; Kuhn, 1970; Latour, 1999). We recognize that there are other epistemic frames one could apply to scientific knowledge building, for example, that it is socially constructed within a community of practitioners (Driver, Leach, Millar, & Scott, 1996; Knorr-Cetina, 1999) or that scientific knowledge is tentative in nature (Kuhn, 1970; Longino, 1990). However, we also believe that classroom discourse and material activity that is framed by the five epistemic features of knowledge above can, in large part, foster these broader understandings of the nature of science, and that no single comprehensible framework for school science investigation can explicitly address all facets of scientific epistemology (see Duschl, 2005, for an extended essay on the relationship between epistemology, school science, and learning).

FOUR REASONS FOR DISCONTENT WITH TSM

Using the above epistemic features of scientific knowledge as a backdrop, we return now to the argument that TSM reinforces unproblematic images of science and oversimplified forms of reasoning. Despite the fact that TSM, when used as an instructional protocol, has allowed many teachers to develop activities that motivate young learners to ask questions, test hypotheses, and work with first-hand data, it misrepresents fundamental intellectual work done by contemporary science. We elaborate below on four specific problems related to the way TSM is used in classrooms.

First, questions are typically provided to students by the curriculum or teacher. This in itself is not an issue, but often students will not be given the resources or experiences to provide a sense-making context within which to understand the purpose of the question (Banilower et al., 2006; Chinn & Malhotra, 2002). In cases where students are given choices of what to study, their questions are often based on what is interesting or doable, but they are rarely grounded in any initial model of what might influence the target phenomenon and in what ways (Carey, Evans, Honda, Jay, & Unger, 1989). As a result, school science investigations can be uninformed and contentless. To illustrate, a few years ago one of our preservice teachers visited a sixth-grade classroom where students were experimenting with plants. She wrote this about her observations:

> Today at school the teacher had students start inquiry experiments. . . They first submitted three questions they could study using plants. The teacher then pointed them towards the ideas that were the most workable. It was really cool to see some of the ideas they came up with. Some students are feeding their plants Coke and others are testing the difference between real and fluorescent light. They are also using dyes and various types of soil—even gravel and sand. One group is testing the effects of music on plants. I was really surprised by the sophistication of their ideas—A lot of questions similar to the ones in our methods class. There was even a plant hanging upside down. It was very cool to see the inquiry method implemented. (Windschitl, 2004, p. 497)

What our preservice teacher may have been enthusiastic about was the sixth-graders' eager engagement in the activities, but what she never questioned was: *Why ask* whether a plant prefers hard rock or a soft drink? Do these students have any reason to believe (or hypothesize) that the physiology of a plant will respond in particular ways to these features in their environment? Because such questions are arbitrary—i.e., make no sense without the context of at least a beginning model for understanding the phenomenon—then any hypotheses emerging from these questions are likely to be little more than poorly informed guesses.[1] This practice reinforces a naive "discovery" worldview in which scientists pose random questions without the framing of any underlying model—tacit or otherwise (Driver et al., 1996; Hodson, 1996).

In more practical terms, this approach to inquiry has no explicit association with scientific concepts, principles, laws, models, etc. This disconnect of inquiry from content is not only antithetical to real science; it can place inquiry on the margins of school curriculum, where accountability pressures for learning subject matter makes inquiry seem like a diversion from getting one's fill of "real science." One of our own student teachers, in fact, recently told us during her practicum that she would "do inquiry on Wednesdays to give the kids a break from content." This artificial separation between investigative science and content has been noted by Roth and Garnier (2007) in their recent analysis of the TIMSS video data

[1] When we have discussed this idea of models being used to formulate questions and hypotheses, some colleagues in the science education community have maintained that scientists or young learners do not always have a model, or even a hypothesis in mind when setting about an investigation—essentially arguing that some studies are purely exploratory. In response, we maintain the postpositivist view that there is no such thing as theory-free observation, and by extension, that no one can hold a motivating interest in asking a question about any natural phenomenon for which one does not have some rudimentary internal sense-making framework (be it a schema, narrative, theory, model, etc.). We recognize, however, that these representations are sometimes so tacit or loosely structured that they are not consciously invoked to develop hypotheses or questions. Take, for example, a first grader who looks out the window at a bird's nest in a tree and wonders how often a pair of birds returns to the nest each day. One's first impression may be that this question, and the interest inherent to it, arises de novo from observation. But if we were to interview the child, we would likely find that she already knows that there are "mother" and "father" birds, that the nest is a "home" constructed by these birds and is not generally not shared by other birds, that they periodically lay eggs, that these eggs hatch and become chicks, that these chicks need regular food, perhaps worms from either the mother or father bird, and that this requires the birds to go out, look for food and to return to the nest periodically. So, the original query by this child may not be very sophisticated, but to say that the child asks such a question without any mental model underlying it, seems irrational.

from science classrooms. In American classrooms in particular, they noted that "almost one-third of the lessons narrowly focused students' attention on performing activities with no attempt on the teachers' part to relate these activities to science ideas" (p. 20).

The second problem with the way TSM is used in classrooms is that it often promotes direct experimentation as the only method of generating data. These experiments feature a comparison between a control group and an experimental group in which a single variable is actively manipulated. In reality there are many ways of testing explanations. Most of them do involve holding constant some factors that might influence the phenomenon of interest so that some alternative explanations can be eliminated (Scott, 2004). But in science fields such as geology, evolutionary biology, natural history, and astronomy, controlled experiments are all but impossible—yet they all use systematic collection of data and coordination of evidence to test new models (Latour, 1999). TSM also suggests that a single set of observations can be definitive in testing an idea (the "critical experiment" as noted by Nadeau & Desautels, 1984), when in fact multiple investigations and bodies of evidence of different kinds are usually brought to bear on assessing a hypothesis.

The third problem is related to the first: because TSM has no provisions for inquirers to develop an initial platform of understanding (a model) to inform their questions and hypotheses, the data are used to characterize only how outcomes are related to conditions (e.g., that small crystals of sugar will dissolve faster in water than large sugar crystals). But underlying explanations based on these analyses (how molecular motion helps break the chemical bonds between molecules of sugar) are rarely addressed (Chinn & Malhotra, 2002; Driver et al., 1996). The idea of scientific explanation is, then, obviated since explanatory argument in actual scientific practice is *not only about patterns in observable relationships* but about how these relationships act as evidence for *why a phenomenon happens in a particular way*. These explanations, of course, frequently involve unobservable processes, events, properties, or structures at a level more fundamental than the observed phenomena themselves (Kosso, 1992; Rouse, 1996; Scott 2004).

In addition to the three previous issues that deal with disconnects between classroom science and real-world science, we hypothesize a fourth problem that is rooted in pedagogical pragmatics—that TSM works too well for teachers. The idea of a self-contained procedure, only nominally linked to conceptual content, with orderly, predictable steps and much of the epistemological complexity stripped away, is actually a useful framework. Because TSM portrays inquiry as a linear process in which each step is a discrete event whose implications are considered only after the previous step is complete (and never to be revisited), this allows an entire investigation to be dispatched within a class period or two. Directing students through well-defined activities in serial fashion satisfies the narrowed intellectual aims for inquiry in many classrooms. In addition, this highly prescribed protocol may be the only form of investigation seen as manageable in today's overcrowded classrooms. TSM, then, becomes a procedure, but *not a way of thinking*. Indeed one can complete the technical aspects of many types of classroom inquiry without knowing the underlying content or being pressed to reason scientifically at all.

If we consider the five epistemic features of scientific knowledge, TSM as commonly practiced does not support intellectual engagement with any of these (Figure 1). One might argue that the testable nature of knowledge is surely invoked in TSM. However, we must ask: *What is it that is being tested or revised?* Using TSM, one often tests the relationship between variables or compares groups on some measure. But in authentic science, this kind of data collection is just one aspect of *testing the viability of an idea*, in the form of a model. When learners talk about data within TSM, the explanatory and conjectural aspects of scientific knowledge are often sidestepped, and in the place of these vital conversations is inserted the vague notion of "stating a conclusion" which, more often than not, is merely

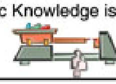
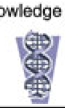
Inquiry as commonly practiced (based on TSM) Goal: *To find patterns in natural phenomena*	Five Epistemic Features of Scientific Knowledge	**Model-based Inquiry** Goal: *To develop defensible explanations of the way the natural world works*
• Hypotheses, often stated as predictions, are tested. • Predictions are not part of larger sense-making theory, so there is nothing to revise.	Scientific Knowledge is Testable  Scientific Knowledge is Revisable 	• Ideas in the form of models are tested and revised. • Accomplished by evaluating hypotheses that make sense within context of a potentially explanatory model.
• Testing culminates in "conclusions" which summarize trends and patterns in the data, but do not include explanations.	Scientific Knowledge is Explanatory 	• Uses patterns in data, other sources of evidence to explain why focal phenomenon happens. • Models talked about as tools for explanation.
• Going "beyond the data" to the theoretical not often a part of classroom inquiries.	Scientific Knowledge is Conjectural 	• Explanations account for observations with underlying, often unobservable causal processes, or structures.
• Models/theories considered to be only an end-product of inquiry, but more often, are not talked about at all.	Scientific Knowledge is Generative 	• Models/theories used to generate plausible hypotheses, new conceptions, new predictions at any point in the inquiry.

Figure 1. Epistemic comparison of classroom inquiry as commonly practiced and model-based inquiry. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

a summary of the data patterns. The discourse of "conclusions" neatly airbrushes out of classroom conversations *why* an investigation turned out the way it did. Finally, with regard to generativity, how can knowledge be seen as a catalyst for new insights if it is portrayed only as the end-product of inquiry?

THE REPRODUCTIVE LEGACY OF THE SCIENTIFIC METHOD

What the Next Generation of Teachers Understands About Scientific Reasoning and Practice

Many who pay attention to how science gets taught in classrooms wonder why research and policy have not had widespread impact on inquiry pedagogies. We believe that the principal reason lies in the deep-seated beliefs about scientific practice that young learners develop over a career of K-16 education. And those individuals who go on to become teachers risk enculturating another generation of learners into familiar but oversimplified images of disciplinary work.

As mentioned earlier, we have conducted several multicase studies, involving more than 60 preservice secondary science teachers. Participants in all our studies held at least an undergraduate degree in an area of science. They had spent hundreds of hours in science laboratories doing all manner of observing, identifying, measuring, data collecting, etc. They had spent years in lecture halls listening to presentations about how science is done, many of these presentations from scientists themselves. It is logical, then, to assume that our participants' postbaccalaureate thinking reflects both how science had been taught to them and how it was likely to be passed down to their own students in the future.

These data we collected were primarily from participants' activities during a secondary science methods course and subsequent student teaching. Our later studies tracked novice teachers into their first year of professional service. Data sources include numerous interviews, participant-generated artifacts from the methods class, recordings of small group conversations, observations in middle school and high school classrooms, and notes from debriefing conversations after teaching episodes.

In our early research, we asked participants during a methods course to conduct an inquiry into a natural phenomenon of interest to them, for which they would collect original data and answer a focal question (Windschitl, 2003; 2004; Windschitl & Thompson, 2004). They were free to use their own images of science to guide these investigations. What we found was a remarkable consistency in how participants thought about and practiced inquiry, and unfortunately, that this consistency reflected an epistemologically impoverished view of science—what we have referred to as a *folk theory* of scientific inquiry. While some facets of this folk theory were superficially congruent with authentic science (e.g., most participants understood that empirical inquiries involve developing questions, designing studies, and collecting and analyzing data; they knew that procedural complications are to be expected in the inquiry process), others indicated a limited view of scientific inquiry (e.g., most participants believed that there is a universally applicable scientific method, although it is not linear, and the ultimate goal of inquiry is to prove or disprove whether a relationship exists between two variables). Most problematic, however, were several facets of this folk theory that misrepresented fundamental principles of scientific inquiry. Many participants believed, for example, that hypotheses function as guesses about outcomes, but are not necessarily part of a larger explanatory framework. They believed that science studies culminate in "conclusions" that merely summarize trends in the data, and many thought that making claims that attempt to link data with unobservable processes was recklessly speculative. Almost entirely absent from participants' inquiry journals and interviews were references to the epistemological bases of inquiry—talk of arguments and claims, alternative explanations, the development of models of natural phenomena, etc. Most of the participants, for example, based their own inquiry questions not on a hypothesized model, but on what seemed novel and somehow testable to them (e.g., bubbling car exhaust through water to see how acidic it became, or exposing seeds to cigarette smoke to see if it would affect the germination process).

Attempts to Scaffold Future Teachers' Thinking About Science Using MBI

We used these early findings to develop scaffolds for participants' thinking about inquiry, emphasizing how the five epistemic aspects of scientific knowledge are embodied in authentic practices (Windschitl & Thompson, 2006; Windschitl et al., 2007). We supported their understanding of the role of models and theory in stimulating productive questions and hypotheses (generativity for the purpose of testing an idea), we required that participants use their data to formulate a causal argument for their outcomes (explanation), that these explanations reference processes or structures that are not directly observable (conjecture), and that they use their findings to modify, support, or refute the model they were testing (revisability). Through these efforts, we have seen participants engage in a form of inquiry that is not only more epistemologically sophisticated than TSM, but just as importantly, is more deeply integrated with science content.

For example, in an early study where participants were *not* given scaffolding, one of preservice teachers, Kyle, conducted an experiment in which he compared the water retention capabilities of different brands of diapers (similar to the common school science activity of testing paper towels). His hypothesis was simply a guess about which brand would hold the most water based on perceived thicknesses of the samples. What followed his careful collection and graphing of the data was a conclusion about which sample held more water, but Kyle made no suggestions about *how or why* water might be retained differently by one diaper versus another. This investigation was systematic and objective, but almost entirely devoid of conceptual content that could explain the outcomes.

A participant in a later study, Jeanne, similarly investigated how fabrics absorb liquids. Members of her preservice cohort, however, were required to "create an initial representation of the phenomenon" they were interested in, to "test a piece of that model" and to culminate their investigations by "connecting empirical findings with an underlying cause." These requirements prompted her to do background reading in both material sciences and the chemistry of liquid–fiber interactions. For her initial model, she proposed a causal system including observable and unobservable processes (how pores in the fabric, molecular structure, and surface area affect absorbency). Jeanne tested for an unobservable interaction that was not common knowledge—the differential affinity of certain liquids to hydrophobic (water repellant) and hydrophilic (water-attracting) ends of fibers. She found supporting evidence for a hypothesis that various fabrics, based on their molecular structure, would absorb different amounts of neutral, acidic, and basic solutions. When she presented her research to peers, she *first* used her experimental design, collection of data, and careful analysis to support her conclusion that indeed various fabrics absorbed differing amounts of liquids (i.e., characterizing the outcomes in terms of the experimental conditions; what we referred to as *method-directed argument*). She *then* used her findings to support her theory about unseen molecular-level interactions that explained the empirical outcomes (what we referred to as *theory-directed argument*).

These two cases illustrate what we mean by testing a hypothesis (Kyle) versus testing an idea (Jeanne). It represents the difference between inquiry as commonly practiced in classrooms—i.e., contentless, little possibility of connections between empirical data and unobservable processes, no press for explanation—and a model-based inquiry approach—i.e., grounded in content, connects empirical data with underlying causes, goes beyond how something happens to why something happens.

Scaffolding participants toward MBI, however, was not as straightforward as we anticipated. Getting preservice teachers to think about testing ideas rather than variables required more than mere prompting. In our first attempt to encourage MBI (Windschitl & Thompson, 2006), we found out, for example, just how many different ways participants could interpret the suggestion to "create a tentative model of some phenomena" and to "test a piece of that model." Some constructed experimental flowcharts rather than representations of the system they were investigating. Others presented a list of statements that all referred to a common phenomenon but were not conceptually integrated with each other, and some framed their models simply as experimental hypotheses. One participant conducted a study of how effective disinfectant sprays were when used on gym equipment. Her model (Figure 2) failed to include any hypothesized representation of the mechanism (bacterial growth patterns, how disinfectant affects the physiology of bacteria) nor did it represent any aspect of the material context of the study (the gym equipment, type of spray)—it was essentially a "black box" representation. This lack of conceptual complexity was repeated, and perhaps resulted in a postinquiry model that was as sparse as the original (Figure 3). The only change in the model was the idea that bacterial resistance needed 3 months to take effect. Only 2 of 21 participants in the cohort, where we first attempted MBI, constructed models of the phenomena with some initial, if rudimentary, explanatory relationships built into them.

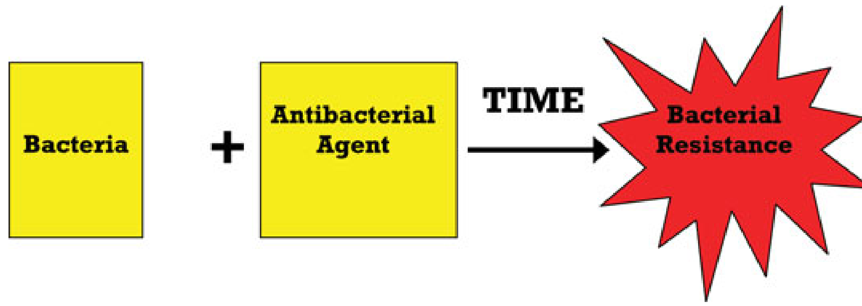


Figure 2. Initial model for inquiry on disinfectant spray use in gyms: shows both content-lean and "black-box" characteristics. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

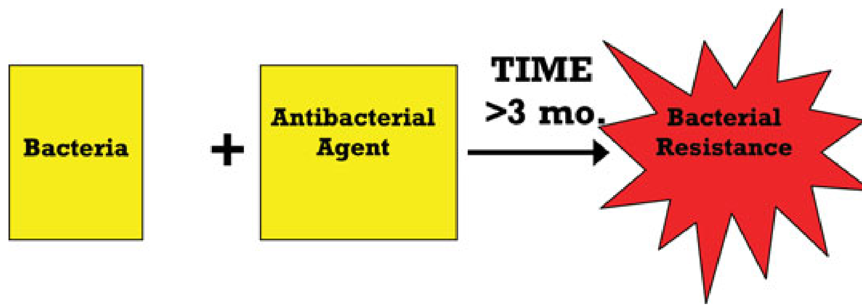


Figure 3. Minimally revised model after inquiry on disinfectant spray use in gyms. Final claim is "bacteria are resistant to spray." [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

From this experience, we developed two lists of model types, one set appropriate for grounding an investigation (e.g., concept maps of the target phenomenon, conceptual drawings) and one set inappropriate (e.g., experimental flowcharts). In the following study then, most participants were able to develop rich representations of the target phenomena and include potentially explanatory processes in them (Windschitl et al., 2007). One participant examined how fruits ripen in response to enzyme activity. The initial model for this study conceivably could have been as much a "black box" as Figure 1. With some background reading, however, this participant represented a far richer set of relationships in her model (Figure 4). Because this participant knew from the outset that she would eventually be asked for explanatory and conjectural arguments, she used her study to elaborate even further on the link between directly detectable influences on ripening (temperature increases, presence of enzymes, etc.) and underlying mechanisms (conversion of methithione and oxygen into ethylene, inducement of gene activity, biochemical effects on plant tissues, etc.). In this case the integration of inquiry practices with conceptual content is clear (Figure 5 is the "postinquiry" model).

Part of scaffolding individuals into MBI became reorienting them to the question "What *am* I testing?" Along these lines, the discourse participants were most unfamiliar with was, surprisingly, that of explanation—the idea that inquiry can culminate in some causal proposition, and as part of that process, that data may be used to infer about theoretical processes. Other researchers have similarly found that preservice science teachers have difficulty constructing arguments involving unobservable hypothesized causes (Lawson, 2002); our own experience suggests they have rarely been pressed to do so, and do not expect it to be "part of the game." A participant in our most recent study, Allison, was

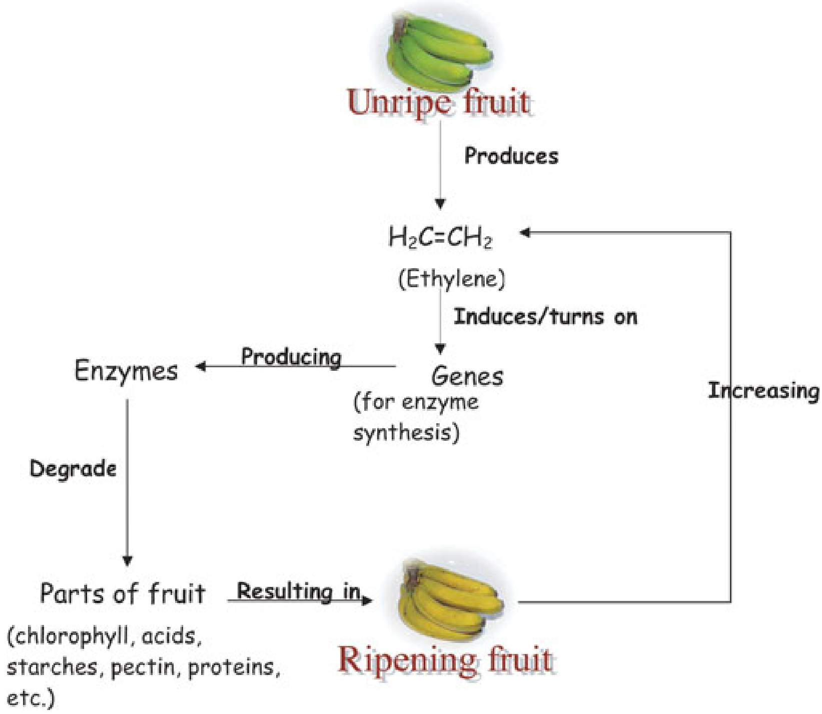


Figure 4. Initial model for inquiry on fruit ripening: shows elements of underlying mechanisms. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

investigating the relationship between the age of trees and their leaf senescence (time at which leaves drop in the fall). In her initial model, she included theoretical processes such as the production of sugars in the leaves and the breakdown of chlorophyll, but had not hypothesized *why* age would have any influence on leaf drop. In an after-class discussion with Allison and other students, we pressed her to consider from her background readings what might influence trees to metabolize differently as they age. Another student suggested that "trees use magnesium or manganese to produce chlorophyll" and that the ability of the trees to take up these minerals could be hindered with age. Allison felt this "opened up" the idea of explanatory mechanism for her, and she eventually incorporated this hypothesis into her model.

In another example, Katlin, who was studying the relationship between enzyme activity in saliva and temperature of foods stated: "I don't have enough connection between my data and the theoretical level. I didn't look at the molecules. I needed, you know, one of those big microscopes that can see at that level" (Windschitl et al., 2007, p. 16). We recognized she was holding to the idea that scientists always directly observe the phenomenon they will eventually make claims about. Katlin also seemed unfamiliar with the discourse of using causal mechanisms to explain experimental outcomes. She later conducted a second pilot study after which she wrote:

> To make matters worse, the data do not appear to address the CAUSE for the relationship between temperature and enzyme activity. This, in my case is the kinetic energy and therefore collision rate of individual molecules. The cause is also the denaturation of enzymes at high temps. Any advice on how to relate the data to the model? [emphasis in original] (p. 16)

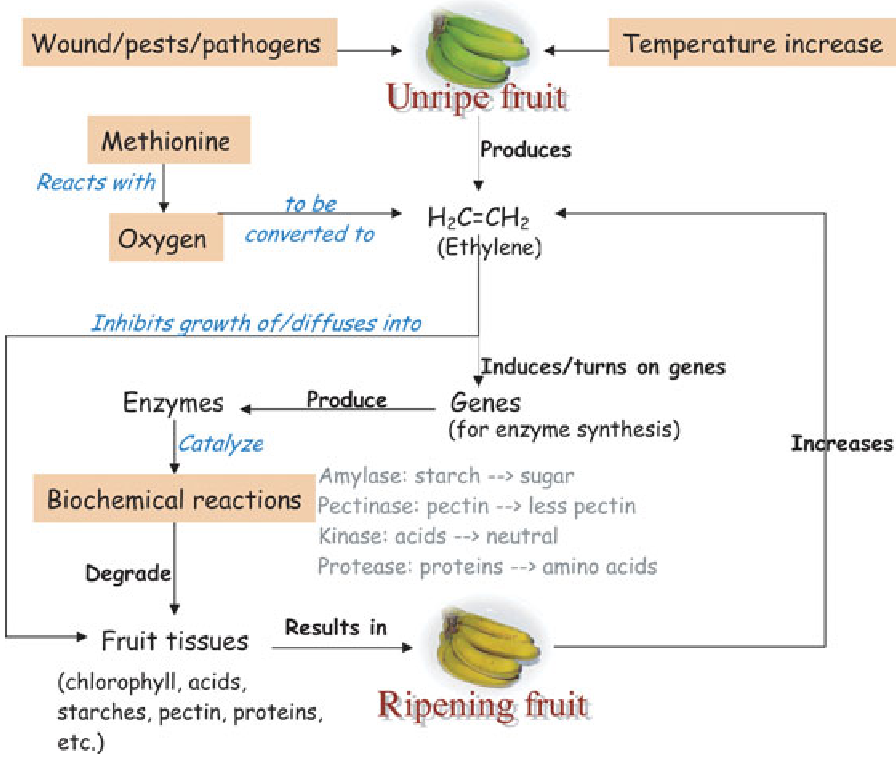


Figure 5. Revised model for inquiry on fruit ripening: rich in conceptual content and explains the phenomenon. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

She then offered to recast her study as a purely descriptive testing of “X against Y”—a regression to a typical school science activity without arguments linking data to underlying explanations. She wrote: “I could present the project as an exploration of the experimental system I used: the interactions of starch, amylase, iodine and temperature, and sharing some of the things I found” (p. 17).

Other participants were more radical disciples of TSM and greeted any alternatives with extreme skepticism. Yvonne, for example, came to the methods course with significant research experience from her precollege and college years. In middle school, she had begun studies on the effects of radiation on fruit flies, had won state and regional science fair competitions, and then had continued this line of research through high school. At the beginning of the course, Yvonne decided that her inquiry project would be an experiment similar to those she had done in middle and high school, but rather than using fruit flies as the objects of her radiation treatments, she used rye grass seeds. She planned to expose the seeds to varying amount of microwave radiation and then test their growth rates after germination. Yvonne maintained that she had used the scientific method for her previous studies, but for the methods class investigation she was forced (by us) to use another—in her words, the “claim and model” method. She wrote in her inquiry journal, “. . . I still feel the scientific method is the best way to approach inquiry for my students.” She gave a summary of how she viewed inquiry and the use of models as distinctly separate enterprises; she included an ironic reference to models as tools:

> From my personal experiences in laboratories I have found that scientists use the scientific method more than anything else. . . Scientists in labs and in the field do use models, but they

> use them *only as tools to understand the phenomena they are studying*. They communicate to other scientists by sharing models they have created based on their research results. [emphasis added] (Windschitl & Thompson, 2006, p. 813)

In a final journal entry, her frustrations with having to think outside the confines of TSM became clear:

> I got really frustrated with having NO choice in whether or not to include models in my independent inquiry project. I think that my project using the regular "scientific method" was much easier to understand than having to have "beginning models" of my system and "revised models" of my system based on my research. I didn't test for any of the factors that I included in my model (which were molecular) so, what's the point of putting it in those terms? [emphasis in original] (p. 813)

Our participants' unfamiliarity and occasional resistance to the idea of using data to make conjectures about processes, events, or structures that are unobservable run counter to the very heart of scientific reasoning. We wondered how they would engage their own students in conversations about evidence—and evidence *for what?* How did they imagine that models of the atom, allelic inheritance, or the core of the earth originated if not from making inferences with data? The most creative and important of all forms of scientific reasoning—theorizing from observation—seemed foreign to many of them.

In our most recent study, we spent months trying to apprentice our preservice teachers into using discourses that connected evidence with theory through the use of scientific argument. However, when asked in a later interview: "What happens in a study after an investigator analyzes the data?" several of our participants relapsed to vague language about how one "draws conclusions" from an investigation rather than talking about scientific argument or explanation. Their responses were similar to their precourse answers, expressing a generally unproblematic view of the relationship between data and any claims made. Approximately half of the 18 participants simply said they would "write up conclusions," "summarize the findings," or "go on to the next question." The language around "claims" and "argument" was not overtly taken up by a number of participants, despite the fact that, in their own inquiry projects, they had struggled with identifying theoretical processes they were making claims about—and, for the most part, doing so successfully.

There is one last piece of data that is a bit more straightforward about helping us understand what needs to change in school science. Across all five of our studies it was clear that participants, during their K-16 school career, lacked opportunities to engage in forms of authentic inquiry in which they posed and answered questions of importance to them, either independently or with the guidance of an instructor. Of the more than 60 participants, only 5 said that they had regular experiences with authentic inquiry. The vast majority could only identify one, perhaps two experiences in which they were allowed to conduct an investigation, and these instances were most often from middle school or high school. The following quote typified their experiences:

> Everything in college and in high school chemistry—everything—was handed to us with a procedure and materials and a purpose for doing it, all laid out like a cookbook. In college it was the same thing in a lab book, and it was all laid down. We never actually had to or got to pursue our own questions or make up our own hypotheses or devise how we would run the experiment. (Windschitl, 2004, p. 501)

While participants could describe material activity—exercises and the like—associated with science coursework, when pressed to identify actual inquiry experiences, most searched their memories in vain. One participant, for example, said as an undergraduate he had spent 6 weeks at a geology field camp doing mapping, but he did not talk about this experience when asked if he had ever "done science." Instead, he referred to a human geography class where he developed and tested his own question via survey about how far people would be willing to commute to school by bicycle rather than by car (Windschitl et al., 2007). Considering the academic histories of these participants, it is hardly a mystery why they find it so challenging or even inadvisable to mentor their own students through meaningful investigations.

INQUIRY 2.0? AN ALTERNATIVE WAY TO THINK ABOUT INVESTIGATIVE SCIENCE

On the basis of our arguments above, we believe there is a strong case that TSM does not reflect authentic science and that it is at least partially responsible for endemic misconceptions in our K-16 educational system about how scientific knowledge is generated and validated. What then, might be an alternative framework for school science investigations? We offer a proposal in the form of MBI (Figure 6), a set of practices that is more closely related to scientific activity and reasoning. We believe this framework can potentially engage learners in discourses embodying the epistemic features of scientific knowledge, press learners to understand content, and can more readily answer the question posed by so many school learners from middle school to undergraduate—"Why are we doing this activity?"

The goal of MBI is to *develop defensible explanations of the way the natural world works*. These words were carefully chosen. In using "explanations," we emphasize that the

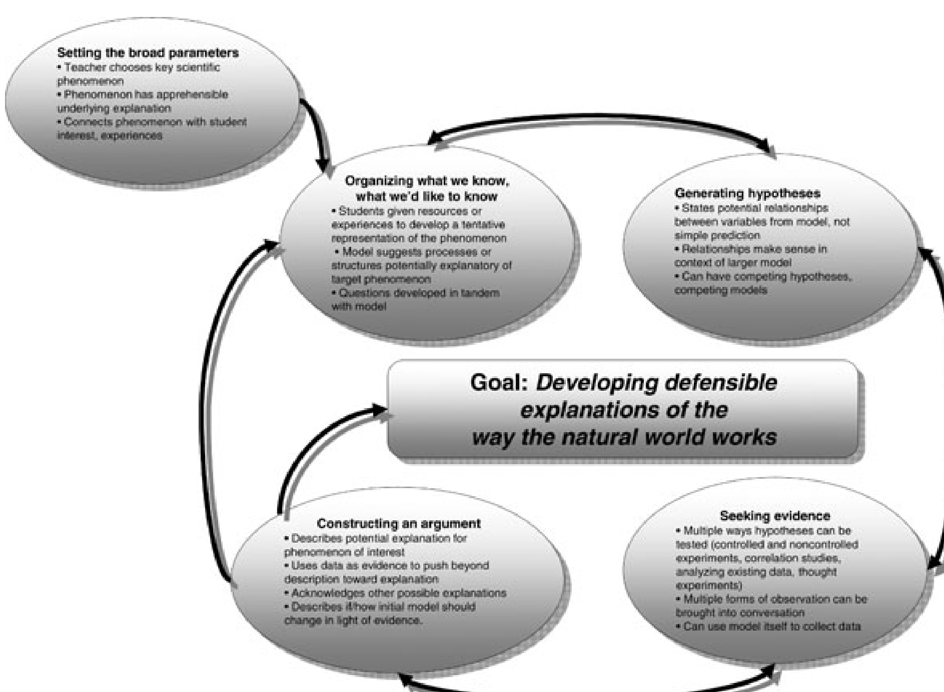


Figure 6. The four interrelated conversations supporting model-based inquiry.

end product should be a statement that helps us understand some aspect of the world at the level of causation. This explanation is embodied in the testable, modifiable representations of scientific ideas, i.e., models. “Defensible” alludes to the use of evidence in creating such an explanation. The “natural world” focuses learners on processes, events, or structures in the general domains of biology, physics, earth science, or chemistry.

Our version of MBI begins with the teacher *setting the general parameters* of what will be studied. This is followed by four kinds of intellectual activity/conversation engaged in by students. We will refer to these simply as “organizing what we know and what we want to know,” “generating a testable hypothesis,” “seeking evidence,” and “constructing a scientific argument.” Some of this activity/conversation is procedural as in TSM, but the emphasis is on purposefully articulating relationships between ideas, evidence, and explanation. In MBI, teachers direct their planning efforts primarily to the intellectual work students will do, rather than focusing on material manipulations. The intellectual discourse of the classroom, in turn, is shaped by the epistemic context of how knowledge evolves within the scientific community. As with other forms of inquiry, MBI can be more or less guided depending on the circumstances. For elementary age learners, we would not expect them to incorporate highly abstract theoretical ideas into models or explanations. We note too that this is not a serial protocol, as a sample case will demonstrate.

Setting the Broad Parameters for MBI

Before the conversations begin, the teacher must set the stage by defining the boundaries of what will be studied, based on student interest and on the scientific importance of an idea. Not all ideas about the natural world are equally important; those that can be *used to explain other phenomena* are more central to science than ideas that are interesting but take you no farther in your understanding of the world. Studying how ideas such as wave motion, inheritance, or chemical equilibrium underlie everyday experiences, for example, will provide students with powerful conceptual tools for understanding much of what they see around them. These constructs are applicable across contexts.

Another concern for the teachers here is that they would have to select a phenomenon with comprehensible causal underpinnings. For example, many middle-grade students know that the shorter the string on a pendulum the briefer its period of oscillation, but could the length of string be considered causal? Even the simplest observable phenomena are often controlled by underlying mechanisms that are complex, contingent, and interactive (see Perkins & Grotzer, 2000, or Hestenes, 1995).

Conversation 1. Organizing What We Know and What We Want to Know. Within this first set of conversations, two principles must be observed. First, investigations should emerge from a motivating interest in some aspect of the natural world and students should be given the resources or experiences to develop a tentative representation (i.e., model) of the phenomenon. This often involves readings, video clips, demonstrations, or even “pre-modeling” laboratory activities. Second, this model should suggest processes, properties, or structures (these may or may not be directly observable) which are potentially explanatory of the target phenomenon.

Developing such representations is important to the entire class because (1) the representations can be shared with others and critiqued, (2) they can help learners see connections between ideas in ways that other representations (such as oral explanations) will not allow, (3) they can be changed as the class (or individual) learns more, and (4) perhaps most importantly, the explicit production of models helps teachers recognize gaps or unique insights in students' thinking that can be addressed before the inquiry can move forward (see Footnote 2 for how expression of models can support student thinking).[2]

Here are a number of "starter" questions that engage students in such conversations (not all can be explored of course, in any one inquiry). The details of these conversations are beyond the scope of this essay, and we recognize that productive dialogic interactions require significant skills on the part of the teacher as well as a sense of where the conversation should be headed.

- What do we already think we know about this situation, process, or event?
- Which of the elements of what we know are based on observations? Inference? Other sources of information?
- Could there be more than one way to represent this situation, event, or process? A conceptual diagram? Pictorial drawing?
- Is our model purely *descriptive* of the situation, process, or event, or does it have parts that try to *explain* what is happening?
- What are we leaving out of our model and why?
- What questions does this model help us ask?
- What additional information do we need to improve our initial model before asking our final inquiry questions?
- How can our questions be framed so that they can be answered by collecting and analyzing data?

Quite often, learners of any age will have difficulty creating even rudimentary inscriptions of the target phenomenon because they lack representational skills (i.e., the ability to translate ideas into drawings or diagrams). Here generous support should be provided by the teacher. In other cases, learners may have the representational skills but lack the personal experiences to generate meaningful models. In this case, MBI may well begin with a hands-on activity so that learners can collectively experience enough of a phenomenon to pique their interest and help them commit to paper a set of ideas and nonarbitrary questions. Inevitably, new questions of interest, insights about relationships, and testable hypotheses arise in conjunction with one another through the development of drawings and diagrams. In practice, these intellectual products and the activities that give rise to them are inseparable. Our sequential listing of questioning, model-building, and developing hypotheses is merely an artifact of communication.

Conversation 2. Generating Testable Hypotheses. The purpose of the "generating testable hypotheses" conversations is to use the model to articulate a possible set of relationships or events for testing. It is important that the hypothesis is testable, but this criterion is secondary to the idea that the hypothesis must make sense within the context of a larger understanding of the phenomenon. Essentially, the hypotheses' place within a larger model helps ensure that it is not arbitrary. This is why we consider this set of conversations critical in MBI, even though some scientists do not explicitly articulate hypotheses for testing.

[2] Externalizing mental models through inscription or dialogue can help learners visualize their ideas in more organized ways and test components of their ideas (Izsak, 2000; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994; White & Fredericksen, 1998). Even mental models, because they can be run "in the mind's eye" to generate predictions and explanations, have been shown to aid both children and adults in experimentation and theory revision (Gentner & Wolff, 2000; Vosniadou, 2002). Models, then, are not simply records of thought, but they shape thinking as well (Olson, 1994).

A common heuristic for structuring school science hypotheses is "If [we observe or manipulate these conditions] then [we should observe this outcome]." However, this results in a prediction about outcomes rather than an expression about how potential relationships within a model might play out in the real world. We advocate for the heuristic to foreground the part of the model being tested: "If we believe that [these relationships within our model] are accurate, then when we [observe or test under these conditions] we should observe [these outcomes]." For example, in a common chemistry activity, a candle is placed upright in a tray of water. The candle is lit and a beaker is placed over it. As the flame dies out, the water level inside the beaker rises. Common competing models for this phenomenon are the following: (1) that combustion of oxygen is primarily responsible for the lowering of air pressure within the beaker, causing outside air pressure to raise the water level (some students "observe" that the water rises approximately 20% in the beaker—the same percentage of room air that oxygen comprises), or (2) that the effect of heating alone causes the air to expand inside the beaker, then contract when the flame goes out, creating the imbalance of air pressure. A student developing a model-grounded hypothesis for option #2 might state it this way: *If* the candle's heating of the air alone causes expansion, followed by contraction of the air inside the flask, *then* when we replace the candle with a noncombusting heating element of the same temperature, *we should observe* that the water level eventually rises to the same level as when the candle was used. This hypothesis situates the test within a larger model, and it implicitly recognizes competing explanations. There can be multiple models and multiple hypotheses that a group of students might test against one another.

Here are some questions that would prompt students to explore the nature of hypotheses and their relationships with models:

- What aspect of our models do we want to test?
- Are there competing hypotheses that could both explain what we are observing?
- When we look at our tentative model and consider the question we want to ask, what would our models predict?
- How can we test our models in a way that generates better descriptions of how this phenomenon happens?
- Can we test our models in a way that helps us understand some process or thing that explains why the models works the way they are predicted to?

Conversation 3. Seeking Evidence. In the "seeking evidence" conversation, questions are posed about how data could be collected to test the model and used to identify patterns or relationships in the observable world. The data will become evidence when it is used to support an argument or explanation (we will see this later in Conversation 4). Data can come from first-hand controlled experimentation, correlational studies, or any other systematic kind of observations. Here are some questions that students need to wrestle with in order to grasp the meaning of "generating evidence" and be able to eventually design their own studies:

- What kind of data would help us test our hypotheses?
- How can we operationalize our variables in ways that will allow us to record unambiguous measurements?
- What will it mean to collect data "systematically"?
- To test our hypothesis, should we observe the phenomenon as it is or actively manipulate some variables while controlling others?

- Can we use a model itself to collect data from (a tabletop watershed model, a bell jar with rubber diaphragm as a model of the lungs, a computer simulation of population genetics), and will that give us valid data?
- When we analyze our data, will we compare groups? Look for correlations between variables? Seek other kinds of patterns and trends?
- What forms of representation (tables, graphs, charts, diagrams, etc.) are most appropriate for the type of data we will collect?
- How should we handle data that are unexpected or run counter to our hypothesis?

Many of these conversations are possible when referring to the designs of existing studies by others, to existing data sets, or even to thought experiments. Talking about evidence in this way, students can understand that this questioning discourse underlies *all* situations in which one would reason about any phenomenon, based on evidence and logic.

Conversation 4. Constructing an Argument. The guiding principle behind this set of conversations is that causal arguments should be constructed that not only attempt to validate the existence of patterns in data but ultimately to support or refute claims about the explanatory processes or entities hypothesized in the original model. Our vision of an authentic argument has four features: first, it describes a potential explanation for the phenomenon of interest; second, it uses the data collected as evidence to support this explanation; third, it acknowledges any other possible explanations that would fit the data; and finally, it describes if and how the initial model of the phenomenon should change in light of the evidence. We specify here that this form of argument is directed toward a supported explanation and can, ideally, employ causal mechanisms that are more fundamental in nature than the tested phenomenon itself. This structure is based on elements from Toulmin's (1958) framework for argument (data, claims, warrants, etc.), which have been successfully adapted by others (e.g., Bell & Linn, 2000; Kelly & Takao, 2002; Sandoval, 2003) for use in educational settings. These adaptations have included two broad discourse goals of (1) articulation of causal claims that explain the phenomenon of interest in which chains of logic cohere sensibly, and (2) using evidence to support or refute these claims, including statements of *how* data relates to claims. As with the other three conversations that make up MBI, we acknowledge that the complexity of classroom dialogue around evidence-based argument can hardly be represented by the following questions, but these questions can serve as entry points into epistemic talk by students.

- Was what our original model predicted consistent with the data we collected?
- How can we go beyond arguing for a simple "cause-and-effect" relationship? For a noncausal correlation between variables?
- Are there intermediate or confounding variables we have overlooked?
- Did we control or at least account for variables that could have influenced the outcomes of the study?
- How do the data help us infer about theoretical (unobservable) events in our model?
- How consistent and coherent is our final explanation for the phenomenon of interest?
- Are there other possible explanations for the data, and if so, how strong is the evidence for these alternatives?
- What did our model fail to predict and why?
- Should our model change in light of the evidence?
- How can our model be applied to other phenomena?

Summary of the Four Conversations. At first glance, it appears that these four sets of conversations should take place in the sequence described here, but this is rarely the case. MBI is an organic process that requires one to constantly revisit previous conversations when new information or insights emerge. For example, in the midst of data collection, even before it is analyzed, scientists often learn things that put them back into conversations about the models and assumptions they are working from. Revisiting previous conversations is the rule rather than the exception.

Of course, no teacher can ask all of these questions during the course of any one inquiry, but every inquiry should include some parts of these key conversations. As students gain experience with guided forms of investigation, they become more competent inquirers by "internalizing" these conversations—eventually asking themselves these questions without prompting from the teacher.

An Example of Model-Based Inquiry

So how might a teacher infuse these critical conversations into an inquiry? We offer an example of an eighth-grade earth science teacher using a guided investigation to help his students understand the relative motions of the earth, moon, and sun. This vignette demonstrates how an instructor can involve students in each of the four conversations in different ways at different points in the inquiry, depending on the needs of learners. It shows how hypotheses can be tested without controlled experiments (or even the testing of covariances among single variables), and it illustrates how inquiry can be used to deepen students' understanding of important concepts in science. The overarching goal of this coordinated set of activities is not to "test variables" but to develop and refine a scientific explanation in the form of a model. This example is drawn from *Inquiry and the National Science Education Standards* (National Research Council, 2000). References to features of the MBI conversations are indicated in italics.

Mr. Gilbert, a middle school teacher, knew that most of his students had difficulty constructing an explanation for the moon's phases that was consistent with their everyday observations. He also knew that a grasp of this phenomenon is important because it demonstrates that objects in the solar system are in regular and predictable motion, and that these motions explain other occurrences such as eclipses, the seasons, and the day/night cycle (*setting the broad parameters*).

Mr. Gilbert began by asking his students what they thought they knew about the moon and listing these ideas on the board (starting the *organizing what we know* conversation). The students called out that "the moon changes shape," "the moon is smaller than the earth," and "people have walked on the moon." Next, he asked about questions they had and wrote these out as well: "Should we try to land on the moon again?," "Why don't eclipses happen more often?," "How wide across is the moon?," and "How often do we get a full moon?"

Mr. Gilbert realized at this point that his students needed more information before they could develop even a beginning model of the behavior of the moon. Using homemade sextants constructed of protractors, straws, and string, he asked students to collect data about the position of the moon in the sky and the moon's shape. He then initiated a conversation around being systematic in collecting and recording this data (a part of the *seeking evidence* conversation; not for testing a model yet, but for creating an initial descriptive model). Some of his students commented that the moon should be viewed at the same time every day. Others added that everyone should follow the same directions for using the sextants.

A few weeks later, when students had made their observations, Mr. Gilbert returned to the moon unit and asked students to display their charts on the wall of the classroom.

Students talked about the patterns they saw in the changing shape of the moon and offered explanations that might account for their data.

Prompted by these charts, some students wanted to know what caused the patterns they had just witnessed. Mr. Gilbert pressed everyone to suggest an *explanatory account* of the observed phenomenon. Some students immediately proposed that the earth's shadow covers different amounts of the moon's surface at various times of the month. Others contended that, as the moon moves through its orbit, we see different sides of the moon illuminated by the sun (starting the *creating a hypothesis* conversation). Mr. Gilbert then asked students to divide into small groups and make a labeled drawing that supported each group's explanation for why the moon changes shape (getting further into the *organizing what we know* conversation). He asked students to talk with one another about how they might use their models to test the two different explanations (extending the *creating a hypothesis* conversation).

The next day, students designed an investigation using globes for the earth, tennis balls for the moon, and an overhead projector for the sun (they participate here in a second round of the *seeking evidence* conversation). Mr. Gilbert circulated among the groups, probing their understandings and focusing their thinking on the relationship between evidence and explanation: "Where would the moon have to be in your model to result in a quarter-moon?" "Show me where the earth's shadow would be." "What evidence do you have that supports your conclusions or causes you to change your mind?" When needed, he asked students to refer back to their chart of the moon's phases and reminded them "A good model will explain that data" (all the questions and prompts here are precursors to the *generating an argument* conversation).

Mr. Gilbert began the next class by asking each group to post their model drawings and invited the rest of the class to examine the results. Most observations seemed to support the explanation that, as the moon moves in orbit around the earth, the amount of the lighted side that can be seen from the earth changes. The students agreed that comparing the order of the phases in their model to the order of moon phases shown on a calendar helped them assess the apparent relationship between the earth, sun, and moon.

One team pointed out that, during the first quarter phase of the moon, the earth's shadow would have to turn at a right angle in order to fall on the moon; they remarked: "That is not the way light and shadows work." Based on such evidence, even the students who proposed the "earth's shadow" model decided to reject it. Mr. Gilbert added a provocative question: "Some of your models predict that an eclipse would happen every month, but we know that doesn't happen. How would we have to change your models so that doesn't happen?" He later asked "Which features of your models work well? Which don't?" (all preceding questions are parts of the *generating an argument* conversation). The students responded that their models still did not do a good job of explaining the height of the moon above the earth's horizon each day, but they did show how the phases of the moon occur. He asked them to do more reading in the library to help them make their models even more consistent with how the moon behaves (returning to the *organizing what we know* conversation).

As a final assessment, Mr. Gilbert asked students to look at the activities the class had completed and record in a summary table all the evidence that supported or refuted the class model of the phases of the moon. While his students completed this task outside of class, Mr. Gilbert used the final 2 days of the unit to explore with his students the classic debates about the earth-centered versus sun-centered models of the solar system and the evidence that Copernicus and Galileo used to support their explanatory model (an extension of the *generating an argument* conversation into historical episodes).

Because Mr. Gilbert knew that the goal of this inquiry was for students to develop defensible explanations of a natural phenomenon, he knew that generating and using evidence would be crucial for his students, but also that this evidence must be applied within the context of a meaningful model for the moon's phases. He coordinated multiple opportunities for students to engage in the four conversations about the relationships between models, evidence, and explanations. Mr. Gilbert scaffolded this talk in judicious ways, providing "seeds" for conversations about the moon's phases, prompting students to pay attention to key aspects of the data in their explanations, and suggesting ways to test their models. And, of course, he pushed his students to go beyond *describing* the phases of the moon and asked them to develop plausible *explanations* for this phenomenon.

It is important to note that MBI, which may seem like the scientific method dressed up in fancy new clothes, is nothing like it in actual practice. It is not formulaic; it is not a set of predetermined experiences; it focuses as much on sense-making discourse as on the material activity; and as it runs its course, it presses learners to engage with all five epistemic features of scientific knowledge.

KEEPING THE LIMITATIONS AND CHALLENGES IN VIEW

We recognize the risks in specifying a framework of activity and discourse that represents the work of science. One could argue that not all scientists explicitly state hypotheses before testing, not all scientists are aware they use models, not all scientists seek explanation in their research, many do not change their models/theories in light of new evidence, and some do not even work empirically.

A number of scholars, for example, have noted cases of scientists who appear to achieve insight without prior hypotheses guiding their observations or their thinking. Simon (2001), for example, cites Planck, the Curies, Roentgen, and Fleming as being driven by curiosity rather than theory when making their most famous discoveries. We do not dispute this. Taking Fleming's case, however, we note that *discovering that* organisms from the genus *Penicillium* somehow causes bacteria to lyse is a far cry from *developing explanations of why* this phenomenon happens. To go beyond discovery, here (as we would want our students to) one needs curiosity *and* a hypothesis grounded in models of chemistry and cellular biology. We note too, as Simon does, that scientists possess extensive and well-integrated background knowledge that makes episodes of revelation seem effortless and unpremeditated. Preceding his "discovery" of electromagnetic induction, for example, Faraday kept a variety of "idea books," records of speculations, common experiments, and theoretical musings (Tweney, 2001). We maintain that the use of models, tacit or otherwise, is part of the preparation for insight. One wonders whether Fleming would have found anything of interest with Roentgen's photographic plates!

Clearly, any instructional prototype designed to support the emulation of disciplinary activity in the classroom—be it science, history, or mathematics—will fail a test of correspondence between the varied practices of the discipline and what can or should be put into practice by educators. The prototype has to represent, in an intellectually honest way, key features of disciplinary activity we want students to understand.

In our conception of MBI, we included hypothesis generation as an explicit component because know that expectations for a reasonable hypothesis can better ensure that learners' intellectual efforts will be directed toward organizing what is known into an initial model. If young learners do not have at least a rudimentary understanding of the theoretical (or conceptual) basis of the phenomena they are studying, the questions they seek to answer become arbitrary, their hypotheses become random guesses, and they have no meaningful context within which to discuss their findings. There is ample room in the "organizing what we know and what we want to know" conversations to engage in prehypothesis observation and testing, but we maintain that at some point, hypotheses must be developed.

Moving on to issues of developmental appropriateness of MBI, adaptations for elementary age learners would be necessary. From their extensive work on both modeling and argumentation, Lehrer and Schauble (2005) recommend that young students begin conversations about models that resemble their target system. They have illustrated this point by having first graders construct a device that "works the way your elbow does" using assorted materials from a hardware store. The initial concerns of their students were around creating a model based on "looks like," which with support eventually developed into a focus on "works like"—attending to relations among and functions of components of the target system. To transition to mathematical modeling of natural phenomena, very young learners would have to develop some grasp of number, measurement, and data. Other concerns include scaffolding children's metarepresentational capabilities (building repertoires of representations, understanding their conventions, purposes), designing systematic ways to observe, and their ability to engage in evidence-based argument. These developmental capabilities have indeed been successfully scaffolded for very young learners (see Magnussen & Palincsar, 2005; Metz, 2000).

The challenges in helping learners become proficient at MBI are not restricted to the elementary levels. As our own data with preservice teachers have shown, even in upper-level classrooms, student are not making sense of science inquiry, they are not familiar enough with the epistemic bases of science, nor do they have the depth of content knowledge to meaningfully answer the question "*Why should we believe* claims about global warming or the benign effects of genetically modified foods?"

Model-based inquiry clearly represents ambitious pedagogy, whether attempted in a fifth-grade classroom or in an undergraduate biology laboratory. Teachers must have deeper understanding of subject matter, broader repertoires of assessment practices, a sense of how to engage students in conversations that push their thinking about content and disciplinary activity, and a long view of coordinating ideas from multiple conversations to bring all learners to a robust understanding of science. There are a number of research reports that describe teachers engaging their students successfully in these types of model-based inquiries (see Schwarz & White, 2005, for a middle school example; and Cartier, 2000, or Wells, Hestenes, & Swackhamer, 1995, for high school).

Typically, however, these teachers receive significant instructional support from researchers who are testing their own versions of MBI with young learners. Without curricula that provide guidance and support for MBI, classroom teachers have to identify for themselves what subject matter ideas may be best explored through MBI, what information resources are available, how to sequence this kind of complex instruction, and when to feel that they have been successful in getting kids to talk and reason scientifically.

CONCLUSIONS

In this essay, we have offered arguments that a new paradigm for framing K-16 science investigations must be developed, and that some version of MBI would be suitable candidate. We do not claim to have invented MBI, rather, we have built upon the work of others in science education, science studies, and the learning sciences. We have also articulated why TSM, as it is commonly practiced, oversimplifies scientific activity and distorts its epistemology and its goals. It also fails to engage learners in deep understanding of content. As a widespread but flawed model of scientific activity, TSM has shaped the thinking of aspiring educators who, in turn, reproduce this mythic narrative for another generation of young learners. We have provided evidence of how difficult it is to "re-engineer" conceptions of investigative science in new teachers, difficult in part because of the confused lexicon we tolerate within the science education community (what do we mean by "explanation," "hypothesis," "evidence," "conclusions," and, of course, "inquiry"?). Some may argue that the shortcomings of TSM are not inherent to its structure, but in how it is interpreted and implemented in school settings. This, however, may be moot because the end result of misguided understandings of science remains the same in either case and any substantive changes in practice would have to originate in new frameworks for inquiry.

Our modest proposal for this alternative framework—*model-based inquiry*—offers a more epistemically congruent representation of how contemporary science is done. From the disciplinary standpoint, the conversations and activities that make up MBI embody the five epistemic features of scientific knowledge: that it is testable, revisable, explanatory, conjectural, and generative. From the instructional standpoint MBI is flexible enough to apply across domains of inquiry and many types of investigations. Importantly too, MBI presses students to engage more deeply with content.

We are under no illusions that one framework, when used as a guide for the design of instruction, can unproblematically effect changes in how science investigations take place in our system of schooling. Because MBI represents ambitious pedagogy, its implementation is dependent upon developing a national corps of highly skilled teachers—yet another major project for science education that remains mired in the status quo. We do not offer MBI as a "final answer," but rather as seed for public conversations from which other, perhaps more pedagogically effective versions of this framework will emerge. We expect that critique, and perhaps criticism will follow our essay, but at the end of the day, can we all agree that this conversation is long overdue?

We wish to thank Dr. Andrea Woody for valuable commentary on earlier versions of this article.

REFERENCES

Abell, S. K., & Smith, D. C. (1994). What is science? Pre-service elementary teachers' conceptions of the nature of science. International Journal of Science Education, 16, 475–487.

Banilower, E., Smith, P. S., Weiss, I. R., & Pasley, J. D. (2006). The status of K-12 science teaching in the United States: Results from a national observation survey. In D. Sunal & E. Wright (Eds.), The impact of the state and national standards on K-12 science teaching (pp. 83–122). Greenwich, CT: Information Age Publishing.

Bauer, H. (1992). Science literacy and the myth of the scientific method. Urbana: University of Illinois Press.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. International Journal of Science Education, 22, 797–817.

Bencze, J. L., & Bowen, G. M. (2001, April). Learner-controlled projects in science teacher education: Planting seeds for revolutionary change. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Bencze, L., & Hodson, D. (1999). Changing practice by changing practice: Toward more authentic science and science curricular development. Journal of Research in Science Teaching, 36(5), 521–539.

Brickhouse, N. (1990). Teacher beliefs about the nature of science and their relationship to classroom practice. Journal of Teacher Education, 41, 471–485.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of 7th grade students' understanding of the construction of scientific knowledge. International Journal of Science Education, 11, 514–529.

Cartier, J. L. (2000). Using a modeling approach to explore scientific epistemology with high school students. Research report 99-1 for the National Center for Improving Student Learning and Achievement in Mathematics and Science. Retrieved November 22, 2005, from http://www.wcer.wisc.edu/ncisla/publications/reports/RR99-1.pdf.

Chinn, C., and Malhotra, B. (2002) Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86, 175–218.

Darden, L. (1991). Theory change in science: Strategies from Mendelian genetics. New York: Oxford University Press.

Dewey, J. (1910). How we think. Boston: DC Heath and Company.

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

Duschl, R. (2005). The high school laboratory experience: Reconsidering the role of evidence, explanation, and the language of science. Paper commissioned by the National Research Council on the Role of the Laboratory in High School Science. Retrieved August 12, 2006, from http://www7.nationalacademies.org/bose/RDuschl_comissioned_paper_71204_HSLabs_Mtg.pdf.

Duschl, R., & Grandy, R. (2008). Reconsidering the character and role of inquiry in school science: Framing the debates. In R. Duschl & R. Grandy (Eds.), Teaching scientific inquiry: Recommendations for research and application (pp. 1–37). Sense Publishers: Rotterdam/Taipei.

Gentner, D., & Wolff, P. (2000). Metaphor and knowledge change. In E. Dietrich & A. Markman (Eds.), Cognitive dynamics: Conceptual change in humans and machines (pp. 295–342). Mahwah, NJ: Erlbaum.

Giere, R. N. (1988). Explaining science: A cognitive approach. Chicago: University of Chicago Press.

Giere, R. N. (1991). Understanding scientific reasoning (3rd ed.). New York: Harcourt Brace Jovanovich.

Hempel, C. G. (1966). Philosophy of natural science. Englewood Cliffs, NJ: Prentice-Hall.

Hestenes, D. (1992). Modeling games in the Newtonian world. American Journal of Physics, 60(8), 732–748.

Hestenes, D. (1995). Modeling software for learning and doing physics. In Carlo Bernardini (Ed.), Thinking physics for teaching (pp. 25–65). New York: Plenum.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. Journal of Curriculum Studies, 28(2), 115–135.

Izsak, A. (2000). Inscribing the winch: Mechanisms by which students develop knowledge structures for representing the physical work with algebra. Journal of the Learning Sciences, 9(1), 31–74.

Kelly, G., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. Science Education, 86(3), 314–342.

Kitcher, P. (1993). The advancement of science: Science without legend, objectivity without illusions. New York: Oxford University Press.

Knorr-Cetina, K. (1999). Epistemic cultures: How sciences make knowledge. Cambridge, MA: Harvard University Press.

Kosso, P (1992). Reading the book of nature: An introduction into the philosophy of science. Cambridge, England: Cambridge University Press.

Kuhn, T. (1970). The structure of scientific revolutions (2nd ed.). Chicago: University of Chicago Press.

Laplante, B. (1997). Teachers' beliefs and instructional strategies in science: Pushing analysis further. Science Education, 81(3), 277–294.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), Representation in scientific practice (pp. 19–68). Cambridge, MA: MIT Press.

Latour, B. (1999). Pandora's hope: Essays on the reality of science studies. Cambridge, MA: Harvard University Press.

Lawson, A. (2002). Sound and faulty arguments generated by pre-service biology teachers when testing hypotheses involving unobservable entities. Journal of Research in Science Teaching, 39(3), 237–252.

Lederman, N. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. Journal of Research in Science Teaching, 29, 331–359.

Lehrer, R., & Schauble, L. (2003). Origins and evolution of model-based reasoning in mathematics and science. In R. Lesh (Ed.), Models and modeling in mathematics teaching, learning, and problem solving (pp. 59–70). Mahwah, NJ: Erlbaum.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. American Educational Research Journal, 41(3), 635–680.

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in elementary grades. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), Understanding mathematics and science matters (pp. 29–53). Mahwah, NJ: Erlbaum.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and scientific literacy: Supporting development in learning in context. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), Handbook of child psychology (6th ed., Vol. 4). Hoboken, NJ: Wiley.

Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T. (2000). Principles for developing thought revealing activities for students and teachers. In A. Kelly & R. Lesh (Eds.), The handbook of research design in mathematics and science education (pp. 591–646). Mahwah, NJ: Erlbaum.

Longino, H. (1990). Science as social knowledge: Values and objectivity in scientific inquiry. Princeton, NJ: Princeton University Press.

Magnussen, S., & Palincsar, A. (2005). Teaching to promote the development of scientific knowledge and reasoning about light at the elementary school level. In M. Donovan & J. Bransford (Eds.), How students learn history, mathematics, and science in the classroom (pp. 421–459). Washington, DC: National Academies Press.

Mellar, H., Bliss, J., Boohan, R., Ogborn, J., & Tompsett, C. (1994). Learning with artificial worlds: Computer based modeling in the curriculum. London: Falmer Press.

Metcalf, S. J., Krajcik, J., & Soloway, E. (2000). Model-it: A design retrospective. In M. Jacobson & R. B. Kozma (Eds.), Innovations in science and mathematics education: Advanced designs for technologies in learning (pp. 77–116). Mahwah, NJ: Erlbaum.

Metz, K. E. (2000). Young children's inquiry in biology: Building the knowledge bases to empower independent inquiry. In J. Minstrell & E. van Zee (Eds.), Inquiring into inquiry in science learning and teaching (pp. 371–404). Washington, DC: American Association for the Advancement of Science.

Nadeau, R., & Desautels, J. (1984). Epistemology and the teaching of science. Ottawa, Canada: Science Council of Canada.

National Research Council. (2000). Inquiry and the National Science Education Standards. Washington, DC: National Academy Press.

Nersessian, N. (2005). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In M. Gorman, R. D. Tweney, D. Gooding, & A. Kincannon (Eds.), Scientific and technological thinking (pp. 17–56). Hillsdale, N J: Erlbaum.

Ochs, E., Jacoby, S., & Gonzales, P. (1994). Interpretive journeys: How physicists talk and travel through graphic space. Configurations, 2(1), 151–171.

Olson, D. R. (1994). The world on paper. Cambridge, England: Cambridge University Press.

Palmquist, B. C., & Finley, F. N. (1997). Pre-service teachers' views of the nature of science during a post-baccalaureate science-teaching program. Journal of Research in Science Teaching, 34(6), 595–615.

Perkins, D.N., & Grotzer, T.A. (2000, April). Models and moves: Focusing on dimensions of causal complexity to achieve deeper scientific understanding. Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.

Pickering, A. (1995). The mangle of practice: Time, agency, and science. Chicago: University of Chicago Press.

Romberg, T., Carpenter, T., & Kwako, J. (2005). Standards based reform and teaching for understanding. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), Understanding mathematics and science matters (pp. 3–26). Mahwah, NJ: Erlbaum.

Roth, K., & Garnier, H. (2007). What science teaching looks like: An international perspective. Educational Leadership, 64(4), 16–23.

Rouse, J. (1987). Knowledge and power: Toward a political philosophy of science. Ithaca, NY: Cornell University Press.

Rudolph, J. L. (2005). Epistemology for the masses: The origins of the scientific method in American schools. History of Education Quarterly, 45, 341–376.

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

Scott, E. (2004). Evolution vs. creationism: An introduction. London: Greenwood Press.

Schwarz, C., & White, B. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. Cognition and Instruction, 23(2), 165–205.

Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: Beliefs and classroom actions. Journal of Research in Science Teaching, 36(8), 930–954.

Simon, H. A. (2001). "Seek and ye shall find": How curiosity engenders discovery. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings (pp. 5–20). Mahwah, NJ: Erlbaum.

Smith, C. L., Maclin, D. Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. Cognition and Instruction, 18(3), 349–422.

Stewart, J., Hafner, R., Johnson, S., & Finkel E. (1992). Science as model-building: Computers and high school genetics. Educational Psychologist, 27, 317–336.

Stewart, J., Passmore, C., Cartier, J. Rudolph, J., & Donovan, S. (2005). Modeling for understanding in science education. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), Understanding mathematics and science matters (pp. 159–184). Mahwah, NJ; Erlbaum.

Solomon, L., Duveen, J., & Scott, L. (1994). Pupils' images of scientific epistemology. International Journal of Science Education, 16, 361–373.

Toulmin, S. (1958). The use of argument. Cambridge, England: Cambridge University Press.

Tweney, R. (2001). Scientific thinking: A cognitive-historical approach. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings (pp. 141–173). Mahwah, NJ: Erlbaum.

Vosniadou, S. (2002). Mental models in conceptual development, (pp. 353–368). In L. Magnani & N. Nersessian (Eds.), Model-based reasoning: Science, technology, values. New York: Kluwer.

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics. American Journal of Physics, 63(7), 606–619.

White, B., & Fredericksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. Cognition and Instruction, 16, 3–118.

Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? Science Education, 87(1), 112–143.

Windschitl, M. (2004). Caught in the cycle of reproducing folk theories of "Inquiry": How pre-service teachers continue the discourse and practices of an atheoretical scientific method. Journal of Research in Science Teaching, 41(5), 481–512.

Windschitl, M., & Thompson, J. (2004, April). Inquiry in pre-service classrooms: Epistemological and methodological aspects. Presented at the Proceedings of the National Association of Research in Science Teaching Conference, Vancouver, British Columbia, Canada.

Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of pre-service instruction on teachers' understandings of model-based inquiry. American Educational Research Journal, 43(4) ,783–835.

Windschitl, M., Thompson, J., & Braaten, M. (April, 2007). How novice science teachers appropriate epistemic disciplinary discourses for use in classrooms. Paper presented at annual meeting of the American Educational Research Association. Chicago, IL.