Developing a Model-Centered Approach to Science Education

Christina V. Schwarz
Michigan State University

Barbara Y. White
University of California at Berkeley
ABSTRACT

We argue that understanding the nature and utility of scientific models and engaging in the process of creating and testing models should be a central focus of science education. To realize this vision, we created and evaluated the Model-Enhanced ThinkerTools Curriculum, which is an inquiry-oriented physics curriculum for middle school students in which they learn about the nature of scientific models and engage in the process of modeling. Key components of our approach include enabling students to create computer models that express their own theories of force and motion, evaluate their models using criteria such as accuracy and plausibility, and engage in discussions about models and the process of modeling. Ten and a half week curricular trials with two teachers in six science classes of an urban middle school indicate that this approach can facilitate a significant improvement in students’ understanding of modeling. Further analysis revealed that the approach was particularly successful in fostering student's understanding of the nature and purpose of models, and less successful in fostering their understanding of model evaluation and revision. The Model-Enhanced ThinkerTools Curriculum also led to significant improvements in inquiry skills and physics knowledge, though there were differential effects of the curriculum for the two teachers, with students of the more experienced teacher gaining more on measures of inquiry and modeling. The substantial one sigma gain on our conceptual physics test was less than that found for the prior, non-model-enhanced version of the ThinkerTools curriculum in which the computer models were all Newtonian. This may indicate a tradeoff between developing a deeper understanding of the nature and purpose of modeling and learning the normative physics. On the other hand, a correlational analyses of pre- and post-test scores suggests that learning about the nature of models can play a role in the
acquisition of inquiry skills and physics knowledge. The paper concludes with a discussion of these results, as well as challenges that remain in enabling students to develop expertise in scientific modeling and inquiry as they develop subject-matter expertise. Finally, we propose refinements that may make this model-centered approach to science education even more effective, and we suggest what further research is needed to advance theories about the role of modeling in the learning of science.
DEVELOPING A MODEL-CENTERED APPROACH TO SCIENCE EDUCATION

Models and the process of modeling are fundamental aspects of science. Advances in technology have served to emphasize their centrality by making model creation and revision even more accessible to scientists. Modeling tools provide scientists with new ways of creating theories, testing ideas, and analyzing data. For example, computational modeling plays a central role in forecasting the weather, in analyzing atomic and molecular structure, in determining the physics of the early universe, and in understanding human reasoning. Because students should understand the processes of science, and because modeling is a fundamental aspect of science that is of great utility to scientists, many science educators have advocated model-centered instruction (Feurzeig & Roberts, 1999). The National Science Standards emphasize this point by stating that “all students should develop an understanding of the nature of science” and that part of this understanding includes the knowledge that “scientists formulate and test their explanations of nature using observation, experiments, and theoretical and mathematical models” (National Science Standards, content standard G, 1996).

Before discussing the benefits and drawbacks of model-centered instruction, we clarify our use of the terms “scientific model” and the “process of scientific modeling.” For the purposes of this paper, we broadly define a scientific model as a set of representations, rules, and reasoning structures that allow one to generate predictions and explanations. Scientific models can range in form from scale models of the solar system, to computer simulations of galaxy collisions, to quantitative laws like \( F = ma \), to qualitative principles such as or “when no forces are acting, an object’s velocity remains the same, because there is nothing to cause it to change.” Models, in
this sense of the term, are viewed as tools for expressing scientific theories in a form that can be used to predict and characterize what will happen as time passes or as events occur. We use the term “scientific modeling” to mean the process used in much of modern science that involves (1) embodying key aspects of theory and data into a model – frequently a computer model, (2) evaluating that model using criteria like accuracy and consistency, and (3) revising that model to accommodate new theoretical ideas or empirical findings. Our model-centered instructional approach is one that places this modeling process explicitly at the center of the learning process.

Benefits of a Model-Centered Approach

Enabling students to engage in modeling has a large number of potential benefits for science education. Model creation and model-based reasoning are core components of both human cognition and scientific inquiry. Students should therefore be involved in a process of creating, testing, revising, and using externalized scientific models that may represent their own internalized mental models (Doerr & Tripp, 1998; Frederiksen & White, 1998; Gilbert, 1995; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994; White, 1993b; White & Frederiksen, 1990). Modeling can help learners to express and externalize their thinking. It can also help them to visualize and test components of their conceptual ideas, which may help them advance their thinking and develop subject matter expertise. Computer modeling can also make some scientific material more accessible and interesting. For example, computer microworlds allow students to have direct access to worlds they cannot experience and give them a sense of owning the dynamics of those worlds (diSessa, 1985; Papert, 1980; White, 1993b). Computer models can also help make complex data more accessible, make the scientific process more dynamic, and allow students to study personally interesting and complex phenomena (Andoloro, Donzelli, &
Furthermore, modeling is increasingly important in society, and many students will need to use computer-modeling technology in their lifetimes (Sabelli, 1994; Tinker, 1993). Modeling approaches that use physical, diagrammatic, or mathematical representations have been shown to enhance student learning (Confrey & Doerr, 1994; Halloun & Hestenes, 1987; Feurzeig & Roberts, 1999; Gobert & Discenna, 1997; Lehrer & Schauble, 2000; Lehrer, Horvath, & Schauble, 1994; Penner, Giles, Lehrer, & Schauble, 1997). Additionally, many model-centered tools and educational software, such as Boxer, Cocoa, Genscope, Explorer, MARS, Model-it, StarLOGO, STELLA, and ThinkerTools, have shown that models can play an important role in promoting subject matter expertise, inquiry skills, and systems thinking (Brand, Rader, Carlone, & Lewis, 1998; Coon, 1988; diSessa, 1985; Horwitz, 1999; Mandinach & Cline, 1993; Mellar, Bliss, Boohan, Ogborn, & Tompsett 1994; Papert, 1980; Raghavan & Glaser, 1995; Resnick, 1999; Richards, Barowy, & Levin, 1992; Snir, Smith, & Grosslight, 1993; Spitulnik, Krajcik, & Soloway, 1999; White, 1993 a&b; Wilensky, 1999). Such modeling tools differ on a variety of dimensions, including whether they focus on problem solving or theory development and whether they engage students in exploring pre-made models (model-exploration) or designing their own models (model-expression) (Mellar et al., 1994). Figure 1 below provides a comparison of modeling software along the expressive/exploratory dimension. In the present study, we explore transforming the ThinkerTools force-and-motion software from primarily a model-exploratory tool to a model-expressive tool.

Insert figure 1 about here
A model-centered approach also has the benefit of enabling students to develop accurate and productive epistemologies of science. If one defines science as a process of model building, this helps students understand that scientific knowledge is a human construct and that models vary in their ability to approximate, explain, and predict real-world phenomena (Gilbert, 1991; Nadeau & Desautels, 1984; Penner et al., 1997; Sherin, DiSessa, & Hammer, 1993; Spitulnik et al., 1999; Stewart et al., 1992). Furthermore, constructing more fruitful epistemological ideas may help students better reason about scientific evidence and better integrate their conceptual knowledge (Driver, Leach, Millar, & Scott, 1996; Gilbert, 1991; Penner et al., 1997; Nadeau, & Desautels, 1984; Songer & Linn; 1991).

Challenges of a Model-Centered Approach

Even though model-centered instruction accurately reflects the purposes and practices of modern science and there is strong evidence that it can help students improve content knowledge and inquiry skills, the challenges presented by such an approach are considerable. In addition to difficulties with teachers’ lack of modeling knowledge (van Driel & Verloop, 1999) and curricular constraints due to state and local standards, some researchers report that students do not always understand the purpose of engaging in the modeling process in a model-centered curriculum (Barowy & Roberts, 1999). There is also ample evidence indicating that students may not understand the nature of models or the process of modeling even when they are engaged in creating and revising models (Carey & Smith, 1993; Grosslight, Unger, Jay, & Smith, 1991; Schwarz, 1998; Schwarz & White, 1998; White & Schwarz, 1999). For example, in a previous
A Model-Centered Approach

study (Schwarz, 1996), we found that students participating in an earlier version of our curriculum could not describe or explain the modeling section of the inquiry process they were using, even though they frequently interacted with a computer model and were creating their own written laws or models. Overall, “there is still a need to examine student understanding and use of models in general and the characteristic knowledge and misunderstandings they hold about models” (Benchmarks for Scientific Literacy, Project 2061).

Some of these problems with model-centered approaches may be due to the fact that such curricula have typically focused primarily on conceptual change. Modeling knowledge is often addressed peripherally and few model-centered approaches have engaged students in direct reflection about the process of scientific modeling for extended periods of time.

Teaching students about the nature of models and the process of modeling has proven to be difficult. Direct efforts at improving modeling knowledge have met with limited success. In prior work, Smith, Snir, and Grosslight (1992) and Wiser, Kipman, and Halkiadakis (1988) developed physics curricula that included meta-conceptual discussions about the nature of models. These curricula had no noticeable effect on students’ level of thinking about models when pre- versus post-instructional interviews were compared.

As a result, we made “meta-modeling knowledge” the primary focus of our 10.5 week curriculum. We hypothesized that engaging students in the modeling process (giving them tools to make their models explicit and testable) and engaging students in direct reflection about this process would be effective in teaching students about scientific modeling. Our extended concentration on model creation and reflection was therefore much more likely to have a
significant impact on students’ modeling knowledge than instructional approaches in which modeling and discussions about models and modeling are more peripheral.

Overview

This paper describes our efforts at reifying the scientific modeling process. To do this, we created, taught, and evaluated the Model-Enhanced ThinkerTools (METT) curriculum in six seventh-grade urban classrooms for approximately 10.5 weeks. This paper also describes the effects of this instructional approach on student learning, and it characterizes the nature of the understanding that students derived from this approach. Specifically, we report on how the METT curriculum and instruction affected students learning about the nature of models and the process of modeling, as well as the effects of this approach on students’ inquiry skills and physics knowledge. We also discuss correlations among these performance measures, which suggest that acquiring modeling knowledge may play an important role in developing students’ inquiry skills and physics knowledge.

Background of the Model-Enhanced ThinkerTools Curriculum

For this research, the authors presuppose a constructivist orientation towards learning. This constructivist orientation influenced our instructional approach in several ways. First, we acknowledge that students require active engagement in model-oriented activities within a classroom research community in order to help them develop an understanding of scientific modeling. As a result, we designed the METT curriculum to include (1) new language about scientific modeling (thereby providing an opportunity for reflection, as in Frederiksen and White, 1998), (2) tools for creating and envisioning scientific models, and (3) opportunities for testing,
debating, and critiquing alternative models. These curricular features are intended to help students develop an understanding of the nature of models and the processes of modeling.

Our work also builds on current theories and evidence about students’ scientific epistemologies, which indicate that students hold what are regarded as naïve views about the nature of science, which influence their learning of science. This naïve epistemology includes a philosophical belief system in which scientific knowledge is seen as absolute truth independent of the investigator’s psychological and social environment (Aikenhead & Ryan, 1992; Ryan & Aikenhead, 1992), a conception of models as copies of reality (Grosslight et al., 1991; Nadeau & Desautels, 1984), and a belief that scientific knowledge does not change (Nadeau & Desautels, 1984), or only changes because new information is discovered (Popper, 1963). Students often use reasoning based on phenomenology, that is, reasoning based on the idea that scientific inquiry involves direct observation and recording of phenomena. Few use model-based reasoning, which is reasoning based on the idea that scientific inquiry involves evaluating theories by constructing models, deducing predictions from the models, and then testing the accuracy of those predictions (Carey & Smith, 1993; Driver et al., 1996; Songer & Linn, 1991). Our approach seeks to enrich students’ epistemologies by engaging them in authentic scientific modeling practices, which include creating, testing, and revising models as well as debating their merits and reflecting on these practices. In this way, our approach attempted to promote a constructivist epistemology and demote the naïve empiricist epistemology that students typically acquire in the science classroom (Nadeau & Desautels, 1984).
RESEARCH QUESTIONS AND THEORETICAL FRAMEWORK

Our primary goal in creating METT (the Model-Enhanced ThinkerTools curriculum) was to design a pedagogical approach that would promote students’ understanding of the nature of models, the process of modeling, the evaluation of models, and the utility of modeling. As a result, our primary research question was, “Can we build model-centered science instruction that will improve students’ understanding of the nature and process of modeling?” We hypothesized that we might be able to obtain an improvement by developing and using a systematic approach to incorporating model-centered activities and orientations. In order to achieve this goal, we incorporated three primary types of modeling activities in the curriculum: (1) students created computer microworlds governed by force-and-motion rules that they choose, (2) students evaluated their model's behavior with criteria such as accuracy and plausibility, and (3) students reflected on the properties of models and the nature of the modeling enterprise. Further, students created and interacted with several different types of models, including diagrams and maps, rule-based computer simulations, and written predictive laws with accompanying causal explanations. Students worked with this modeling approach over a significant period of time (about 10.5 weeks).

Our second research question was, “What effect does a model-centered curriculum, which is aimed at improving students’ understanding of models and the process of modeling, have on students’ inquiry skills and their conceptual physics knowledge?” We hypothesized that an understanding of the nature of models and modeling, including knowing about model forms as well as their creation, evaluation, and utility, would play an important role in learning about scientific inquiry and in developing a conceptual understanding of scientific phenomena like force
and motion. A related third research question that we addressed is, "How is the development of modeling knowledge related to the development of science knowledge and inquiry skills?"

With regard to the acquisition of science knowledge, we hypothesized that knowing about the nature and form of the scientific product (the epistemic forms and models) as well as how that product is generated (the epistemic game or modeling process) might help students better produce and use those products (Carey & Smith, 1993, Collins & Ferguson, 1993; Gobert & Discenna, 1997; Schwarz, 1998; Schwarz, 2002b). In other words, understanding that models are representations that can predict and explain may help students construct models that are predictive and explanatory as well as use models to predict and explain. For example, Smith, Snir, & Raz (2002) found that enabling students to understand that models can be useful explanatory tools helps students use models in reasoning about the phenomena they were studying. In addition, knowing that models are constructed artifacts, which are not identical to reality and therefore have limitations in their ability to represent the system that they model, may help students understand that they need to be careful not to internalize all aspects of models (Schwarz, 2002a). Finally, understanding that models should be evaluated using criteria such as accuracy may help students evaluate their own conceptual models with such criteria.

With regard to the development of inquiry skills, we hypothesized that modeling knowledge may help students better learn to engage in scientific inquiry, particularly since modeling is a key component of inquiry. Knowing the forms and purpose of models may help students have a clearer goal of the inquiry process. This in turn may help students form more specific hypotheses, come up with experimental designs that generate data which support or
disprove a certain model, and summarize their conclusions in the form of a model such as a predictive or explanatory rule.

THE MODEL-ENHANCED THINKERTOOLS (METT) APPROACH

The METT curriculum is based on preliminary studies by the first author (Schwarz, 1996) and earlier work from the ThinkerTools Inquiry Project (White & Frederiksen, 1998; White and Horwitz, 1988; White, 1984 & 1993a).

The instructional framework for the METT intervention incorporates pedagogical principles similar to those espoused in the ThinkerTools Inquiry Curriculum (White & Frederiksen, 1998). Some of these pedagogical aspects include the instructional techniques of modeling, scaffolding, and fading from cognitive apprenticeship (Collins, Brown, & Newman, 1989), enculturating students into a scientific community with tools (such as computer simulation models) and language (such as introducing the term ‘model’) from social constructivism (Brown & Campione, 1996; Driver, Asoko, Leach, Mortimer, & Scott, 1994; Palincsar, 1998; Vygotsky, 1978), and incorporating practices of the scientific community (such as scientific inquiry, modeling, and scientific debates) from situated cognition (Brown, Collins, & Duguid, 1989).

The METT approach incorporates multiple reflective components to create opportunities for the elaboration and refinement of knowledge structures. Metacognition and reflection have been found to be critical components of these learning processes and have been shown to greatly enhance learning outcomes (Brown & Campione, 1996; Gunstone, 1991; Kuhn, 1989; White & Frederiksen, 1998). Students in METT engage in a variety of reflective activities:
They reflect on the nature of models after reading various passages about the nature and utility of models, they reflect on the accuracy and plausibility of their models with respect to their own real-world data, and they reflect on the quality of their scientific reasoning and teamwork (plus other self-assessment categories used in the prior ThinkerTools curriculum – see White and Frederiksen, 1998).

Before describing the implementation of our METT curriculum and how it differed from the prior ThinkerTools curriculum, we first describe the Model-Design software, which is the primary modeling tool used in the curriculum.

The Model-Design Software

The authors, in collaboration with Christopher Schneider and John Frederiksen, designed the model-enhanced version of the ThinkerTools software (Schneider, 1999). In developing the METT software, our goal was to enable students to easily create and compare alternative models of force-and-motion phenomena, without having to program. As a consequence, METT is a more domain-specific and limited modeling tool than StarLOGO or BOXER or even Model-it. It allows student to easily create microworlds that obey various alternative laws of motion and thus behave in very different ways. This type of model-creation scaffolding can have pedagogically utility in helping students focus on particular conceptual ideas about the subject matter as well as ideas about the process of modeling.

To enable students to investigate the implications of their own ideas about force and motion, the METT software allows users to change the “laws” of motion that control the simulation software. Students can do this by choosing from a set of alternative rules of motion, the rule that corresponds most closely to the physical laws they derived from their real-world
experiments. Each set of alternative rules, such as the set for no-friction, or sliding friction, or fluid friction, includes Newtonian and non-Newtonian rules. (See Figure 2 for screen shots of two sets of rule choices offered by the software.)

For example, if the data from their real-world experimentation indicated that reducing friction caused an object in motion to travel at approximately a constant speed over time, students might then choose the Newtonian model “constant speed.” After choosing such a rule from this set of three (or four) qualitative or semi-quantitative rules, the students are then asked to compose a mechanistic, causal explanation to justify their choice by responding to the prompt, “I think this is true because: …” (such as, “because there is nothing to make it slow down”). Finally, students run the simulation in order to see the consequences of their chosen rule. (See Figure 3)

As Figure 3 indicates, running the simulation that is associated with any of the three rules causes the object to behave according to that rule. A student can create more complex microworlds with objects, walls and targets and can turn on forces like friction or gravity. Any new microworld that the students create will behave according to the rules they have chosen to govern that microworld. For more information about the ThinkerTools software and the representations it
A Model-Centered Approach employs, see White and Frederiksen (1998). Also, see Schwarz (1998) and Schneider (1998) for further discussion about the design and revision of the modeling software.

Allowing students to develop their own models, based on data from their real-world experiments, and then to explore and revise those models, based on further thought and experimentation, addresses the essence of scientific modeling as a theory building enterprise. The computer modeling activities that we developed were meant to help students learn about the nature of models (a model can be as simple as a rule that allows someone to predict and explain a phenomena; models are not necessarily real or correct, but good ones are better estimates of a phenomenon; there are multiple models for the same phenomenon), the nature of modeling (modeling involves embodying key parts of a theory into rules and representations), the evaluation of models (models can be assessed using criteria like accuracy and plausibility) and the utility of modeling (models are useful for envisioning or testing a theory and for deriving consequences that may be useful for solving problems).

In addition to allowing students to modify the simulation by choosing the rules that it follows, the software also allows students to select “Newtonian Model-Design” and see the simulation run according to Newton’s laws of motion. Students were encouraged to compare the behavior of their own models to the Newtonian model. In this way, we attempted to foster the idea that scientific inquiry is a process of comparing and testing competing models and that one uses criteria such as accuracy and plausibility to make these comparisons.

Instructional Implementation of the METT Approach

Students in the METT curricular approach conduct their research on force and motion by following the scientific inquiry cycle shown in Figure 4. This cycle is repeated four times as
students explore the topics of one-dimensional motion, with and without friction (modules 1 and 2), two-dimensional motion (module 3), and either gravity, mass, or gas/fluid resistance (module 4, which is a research project). The following text describes the classroom implementation of these topics within the inquiry cycle, primarily using the first module as an example.

---

**Introduction and Motivation**

At the beginning of the curriculum, the first author or the teacher explained to the students that they would be learning about scientific modeling and that modeling is an important part of science. Students then completed a series of pre-tests and discussed a videotape of modern uses of computer simulation models, which further introduced them to the idea of models and modeling. The videotape included a segment of a computer simulated tornado storm, a simulation of two galaxies colliding, some impulse-based simulations of objects moving on surfaces (Mirtich & Canney, 1994), and a short clip of the video animation movie, “Toy Story.”

Watching and discussing the videotape of modern computer simulation models addressed several goals. It provided an opportunity to discuss the nature and utility of computer simulations. For example, the video images demonstrated how models could be useful for visualizing phenomena and for performing difficult experiments or experiments not otherwise possible. The videotape also showed students various types of computer models and motivated them to understand scientific modeling. Months after the curriculum and instruction ended, many students commented that this videotape was exceptionally memorable and interesting. The
following edited transcript is an example of the dialog that occurred after students watched the introductory videotape. The transcript is meant to illustrate how some of the modeling language and concepts were introduced to students.

CS: … [Now that we have watched this videotape,] what do you all think computer models and models are in general? Does anybody have any ideas? …

S1: Models and computer models are … a simulation to make it look like [the object] … so that you make sure it work[s] right, and that it be in the right place.

CS: Great. … So one idea is that it’s something that you want to look similar to the original object so you can see if whatever it is that you’re testing actually works. … Do any of you have other ideas for what they might be? …

S2: Um, … since they are similar to the actual object, they’re um, supposed to help you predict things.

CS: Exactly. So like the thunderstorm video that we saw yesterday was a good example of that, because ideally in the future, we’ll have such great computer models that we’ll be able to predict when a tornado is going to happen or a big, huge thunderstorm, or a hurricane, or whatever it might be. Um, what do you think they’re used for other than prediction? …

S2: Um, to see whether theories are true.

CS: Okay. Can you explain a little more about what you mean by theory?

S2: Um, well I think that a theory is kind of like what your prediction would be. A lot of things you don’t know, and so we’re just guessing. Um, and maybe, it’s good to test someone else’s theory ….
CS: ... So, computer models are really good for testing theories. …For the galaxy collision example, for instance, there are some astronomers who think that particular galaxies collide and form other galaxies. But they need to be able to test that and to see. So they run computer simulations or computer models and see what the outcome looks like. And then they go off with their telescopes and they see if they can observe anything like that.

We note from the previous transcript that some students are already showing some sophisticated understanding about the nature and purpose of models. This kind of discussion can potentially provide a fruitful introduction to modeling, and it can provide a foundation for students who have had less experience thinking about ideas related to modeling.

Once the classes had finished watching and discussing the modeling videotape, students received the first ThinkerTools module. The first module further introduced the curriculum and the scientific inquiry cycle, and it discussed why students should study force and motion.

**Question and Hypothesize**

The first phase of the inquiry cycle begins with an interesting scientific question. In METT, students are given the question and are asked to make alternative predictions about possible answers to that question. As in conducting research in science, questioning and hypothesizing are important steps in creating a scientific model and are important precursors to developing an experimental design. In module one of METT, students were asked to think about why an object moving on a surface slows down and eventually stops. Students read several "predictive questions" and formulated hypotheses. For example, students were asked the predictive question, “Imagine that an object has been hit and is moving on a rough surface like a
carpet. How do you think a large amount of friction (caused by the object moving over a rough surface) affects the speed of the object as time goes by? How do you think friction causes this to happen?” The following example from a student's research book is a common response, “The object would probably slow down and stop after a while.” Such responses are often accompanied by an explanation like, “It’s rough and has things sticking out that would make the object have to run over and slow down.” Once students answered this predictive question, they were asked to form hypotheses about what happens to the speed of an object that has been hit on a very smooth surface, and gave responses such as “it slows it down, but not as much.” Then they were asked to form hypotheses about what would happen on a surface that is so smooth that there is absolutely no friction, and gave responses such as “it would stay the same until the surface stopped” or “[it] slows down and stops, but takes even longer” or “[it would] pick up speed or stay the same velocity”. Students shared their predictions with the whole class. The aim was for them to learn that there are alternative theories that can predict and explain the motion of an object.

Investigate

In the investigation phase of the inquiry cycle, students explored possible answers to their research questions by designing and conducting real-world experiments. Obtaining good experimental data was critical for METT students, because their computer models would be grounded in rules derived from their empirical findings. For example, students in the first module gave a smooth plastic puck an impulse (a standard-sized hit given with a mallet) and measured the object’s speed over various distances. This enabled them to see whether the speed increased, decreased or stayed the same. The students then repeated the experiment in a reduced-friction-
environment. These classroom experiments were conducted in teams with each student playing a different role. For example, one student might give the puck a ‘standard’ bonk with the mallet. Another student might measure the amount of time the puck took to travel two meters, and yet another would record the data. The experiments necessitated teamwork, as they frequently required coordination of several pieces of equipment.

This experimental phase of METT diverged significantly from the prior version of the ThinkerTools curriculum. Students in the prior version supplemented their real-world experiments with experiments using the ThinkerTools Newtonian microworlds. In other words, they would explore various environments and games in the ThinkerTools Newtonian microworlds to help them discover the Newtonian physics (White, 1984; 1993a; White & Frederiksen, 1998; White & Horwitz, 1988). In this sense, they were using the software in model-exploratory mode rather than a more model-expressive mode. In the METT version, however, the computer model was used in more of a model-expressive mode: Students’ used the software to model their conceptual ideas, not to discover the normative physics.

**Analyze**

In the analysis phase of the inquiry cycle, the students analyzed the data from their real world experiments by looking for patterns. Again, this step was critical in that students needed to carefully interpret their data so they could form accurate models. For example, in the first module, students determined the differences in speed between the first and the second meter that the puck traveled, with different amounts of friction, to see whether the speed increased, decreased or stayed the same. For much of the curriculum, the entire class completed the analysis section together and discussed the findings from their experiments.
The model phase of the inquiry cycle had two purposes – to help students reify their alternative models, based on the empirical evidence from their investigations, and to help students reflect on the nature of models. In the beginning of this inquiry phase, students formed a predictive law to summarize their findings as well as causal models to explain them. The following excerpt from a student's journal is a typical written law from the first module, “Friction causes objects in motion to slow down and eventually stop, because the rubbing takes away the speed.”

Next, we introduced the computer modeling activity by informing students that they would be incorporating their law into a computer model so that they could see the behavior it predicts, which is an important step for science students and scientists alike. We explained that students would be comparing the behavior predicted by the tentative rule they created (which characterized their experimental results) to the behavior of the real world to see if they matched. The computer modeling software would allow them to visualize and test their modeling rule.

For the first and second modules in the curriculum, the teacher then reviewed the computer instructions with the students in order for them to use the computer more productively. After the instructions, the classes were divided into two groups that differed in the order of two activities: (a) computer modeling and (b) reading about the nature of models and modeling. One group of students began with the computer modeling activity, and then did the reading activity. The second group began by reading passages about the nature of models and modeling, and followed this with the modeling activity. The activities were conducted this way because there were only enough computers for half the class to work on them at any given time.
As we previously described, the pairs of students on the computer chose the computer modeling rule that most closely corresponded to the rule that best fit their experimental data. Once students at the computer had chosen among the three or four qualitative or semi-quantitative computer rules, they ran the associated computer simulation in order to see the model’s consequences and compared it to the Newtonian version. To give a better sense of the student interaction with the computer models, we present a portion of a transcript of three male students working on their final project in order to address the question, “How does the mass of an object affect what an impulse does to its motion?” When the students were at the computer during the modeling phase of the inquiry cycle, they said:

S1: Well, what rule do we choose? [He reads from the directions for the project] “Find the model design rule on the computer that best fits your conclusions.” … Which one [should we pick]? Lighter goes faster?

S2: Yeah, put that one

S3: Yeah, yeah, that one

S1: Why? … The heavier object goes slower because it’s heaviest

S2: It’s heaviest and it puts the lighter object

S1: Rubs against the surface. Yeah.

S2: See if we’re correct. [They run the simulation] DANG! It do.

S1: Maybe it is!

S3: … So the lighter do go fast. Now go. ... [And they move on to trying out different modeling rules]
Pairs of students not working on the computer were engaged in reading and reflecting on several reading passages about models and modeling in the first two modules. These sections, entitled “collaborative thinking about models” included three passages about what a scientific model is, how the ThinkerTools computer program works, and the utility of computer models. For example, in the passage on what a model is, students compared and contrasted three different maps of the area around their school in order to discuss advantages and disadvantages of different representations. Students read the passages and summarized the content to each other in order to reflect about the nature of modeling. The following portion of a transcript illustrates a conversation between two students about what a model is. It is taken from the ‘collaborative reflections about models’ section in the first module of the curriculum.

S1: (reading the passage aloud): “What is a scientific model? In science, we have a special meaning for the word model. A model is a set of representations and rules that allow you to predict and explain what will happen. For example, the rule ‘roughly every twenty four hours, the sun rises in the east because the earth rotates on its axis’ can be thought of as a model. A model can also be a description of an object or system that attempts to explain all of its known behaviors.”

S2: What are some models, what do you think are some of the models that we see every day? Like, maps, globes? …

S1: [S walks to pictures or diagrams on the walls and points them out.] That’s a model, that’s a model.

S2: What about, like, like the TV when you watch that?
S1: That’s a model. It’s a model showing, it’s something that’s really happening. But, it’s not the real thing. …

S2: Okay, so.

S1: The computer is a model.

S2: The computer simulations are models.

S1: Uh, huh. [back to reading the passage] Okay, “Creating models is a critical part of doing science. For some scientists, creating models is their main goal. One way to create scientific models is to use the computer. For example, the computer program in ThinkerTools allows you to create the models that describe and predict the motion of objects. Creating models with the ThinkerTools software can help us better analyze, clarify, and express our theories about the behavior of objects and systems.”

S2: Okay. My turn. I want to read this one. “Models can also be useful to develop and test ideas. For example, different scientists can have different theories, which can be programmed into computer models. Scientists can then play with these models to see implications of their theories and test their theories. Here are some examples of computer models: Computer models that astronomers use to study how galaxies interact and collide.”

S1: “Computer models that take into account the amount of pollution in our atmosphere and how that affects the weather patterns in our air.”

S2: “Computer models of atoms and molecules.”
S1: “In your own words, describe what you think a scientific models is.” … It, a sci, a scientific model is a small, is a small. A small … representation of something that is going to be built into a bigger thing to test your theories and test stuff on it. …

S2: It doesn’t necessarily have to be smaller. I mean

S1: It can’t be huge if you’re going to build another thing out of it. ‘Cuz I mean, you’re not going to make a huge model, and then make a real

S2: A scientific model, it could be computer simulations.

S1: [Writing down and talking aloud] Plans, toys, smaller houses, doll house.

S2: Scientific model.

S1: That is a scientific model if you’re going to make a bigger house out it. It’s going to make…

S2: I don’t think they use a doll house for [science].

S1: Well, they can build a doll house … but still… build it into a bigger house. Mrs. H.’s husband did that. Anyway. So, TV’s?

S2: No. [A] scientific model. … Like what would they use? Like scientists they figure out, is a. Like what they said. Atoms and stuff. …

S1: Solar system models.

S2: Yeah. Like that. Wait, wait, wait. We’re doing this wrong. It says, what do you think a scientific model is.

S1: Computer simulations, plans, toys, smaller houses, a doll house, solar system models, etc. …[reading directions]“Give two more examples of scientific models other than the ones already mentioned.”
S2: Um globe.

S1: A globe, yeah.

S2: What about satellite pictures?

S1: Yeah. … [she reads] “What makes your examples scientific models?” They all show

S2: They show, what? Things that scientists use?

S1: Small replicas

S2: That scientists use.

This dialog illustrates some of the difficulties that students have in understanding the idea of a scientific model, including a focus on the size of a model and a difficulty in understanding the difference between a representation and something that is not ‘real.’ The dialog also shows how the students are trying to make sense of the information and how they are beginning to move towards thinking about models as abstract objects that are useful for testing theories.

In a different class discussion, related to a reading passage about the utility of models, students raised many points, including the ideas that scientific models can be useful for getting more information, having fun, playing games, learning, designing and making things, doing research, making predictions (about earthquakes), testing the effectiveness of seatbelts in cars, and seeing how to successfully launch a rocket.

Finally, at the end of the model section of the inquiry cycle, after all students have finished both the computer modeling and reading activities, teachers are asked to engage their students in a whole class discussion about the computer modeling rules, and to ask students to share which modeling rules they thought were best and why. We note that students in the prior version of the ThinkerTools curriculum did not interact with the Model-Design aspect of the
software in the modeling phase of the inquiry cycle. Instead, during the modeling phase of the inquiry cycle, they created laws and causal models to summarize the results from their real-world and computer-model experiments, and then the class tried to reach a consensus about which were the best laws. In this sense, students were creating modeling laws (engaging in model-expression), but their laws were expressed as written statements rather than as computer models. Additionally, students in the prior version of the ThinkerTools curriculum did not engage in reflection about the nature of models or the process of modeling.

**Evaluate**

In evaluation, the last stage of the inquiry cycle, METT students evaluated their models with respect to modeling criteria such as accuracy and plausibility. This step was critical in helping students understand that models must be evaluated to establish their validity and to improve them. The evaluation activity was introduced in the first module by discussing these points with students, and by asking them what criteria they would use to evaluate a model. This activity served to introduce the four main criteria for characterizing good models that are used within the METT curriculum: accuracy, plausible mechanism, utility, and consistency.

Students evaluated their models by choosing a score between one and five to represent how well their model satisfied each evaluation criterion. Students also justified their choice with a written response as well as an oral response presented to other students in class discussion. For an example of model evaluation instructions, see Figure 5 from the research book. The model mentioned in the directions refers to the computer modeling rule students’ chose to represent their own conceptual model.
For example, one student gave her model a five for accuracy, and stated in her research book, “the real world experiment and the computer model was about the same.” Another student gave her model a ‘4’ for plausible mechanism and stated, “I gave our model this rating because we don’t actually know if an object would keep going [forever], but it seems right.”

We hypothesized that having students evaluate their models with criteria would help them understand that some models are better than others with respect to the evaluation criteria, and that there are multiple scientific models, none of which represents absolute reality. We also note that, at every stage of the inquiry cycle, students also assessed their own work according to other criteria from the prior ThinkerTools curriculum such as being inventive, reasoning carefully, and teamwork (White & Frederiksen, 1998).

To illustrate student talk during model evaluation, we present a brief excerpt of a student discussing his model during his final project presentation. In his project presentation, the student, JP, tells the class that the results from his real world data (that an object slows down when dropped through a thick fluid like honey) conflicted with one of the non-Newtonian simulations that he believed was correct it shows that, if an object keeps slowing down, it will eventually stop. He stated,

We estimated the rule before we went onto the computer, and our rule was that, um, when an object is affected by friction, it slows down. So we went on the computer, and we tried out that [Model-Design rule]. That [Model-Design rule] was um, that the object slows down. … And the object slowed down so much that it didn’t even go all the way,
which wouldn’t happen in real life. So we tried um, … the one where you could use it at a constant speed, and that seemed to work right, which puzzled us a little bit because that’s not what our data showed. So, we were a little inconsistent there. So, on our ratings at the end, our self-assessment, that kind of messed us up ‘cause… we had to be careful. We weren’t exactly all-too careful, obviously, and the data showed that we messed up.

Once students evaluated their models with criteria, they worked on several model application questions, which asked students to apply their model to practical situations. See figure 6 for a sample application question.

Some sample student responses to this question include “I picked A because even if she hits it with less energy, the puck will reach the other end because there’s no friction to stop it.” And, “I think my answer is right because the rink doesn’t have no friction (sic) so the puck shouldn’t slow down.” Students worked on these application questions individually and then shared their reasoning and answers in a class discussion. The application questions helped them apply and explore their conceptual models. Discussing the application questions then led students to evaluate the limitations of their models, which helped them understand the need for including additional complexity in their subsequent models. Raising limitations provided motivation for conducting investigations in the next module of the curriculum, such as adding in another
dimension to their analysis of force and motion phenomena (2-dimensional motion) or including other forces such as gravity.

Before moving on to the next module and inquiry cycle, students in the final portion of the evaluation stage compared and contrasted their model to other students’ models in class debates (cf., Bell & Linn, 2000; Koslowski, 1996; Kuhn, 1993). This debate activity motivated students to carefully choose and evaluate their models, and the debates served to make the classroom process similar to the evaluation method used within the scientific community. For example, in module one where students are asked to debate the three models of motion (objects slow down, speed up, or remain at a constant speed when there are no forces like friction acting on them), students used a host of tactics. They appealed to logic, “How is it [an object in motion] going to stay the same?” to which a student responded with “How is it going to stop?” Students appealed to practical arguments such as “There is no such thing as a frictionless surface!” to which another student responded, “Pretend!” Others appealed to evidence stating, “Prove it!” which was followed by a discussion about pucks on surfaces and spaceships in outer space. Students also appealed to the computer model to help them prove their point, although some students raised the issue that “you programmed yours to do that.” And “It’ll do whatever you want it to do!” The argument finally ended when a student stated, “Just listen to the rule. … It’s a good rule.”

The evaluation phase of the prior version of the ThinkerTools curriculum differed significantly from the METT version. Students in the prior ThinkerTools curriculum completed application questions that asked them to apply their conceptual models, which they had written in the form of laws, to predict what would happen in the ThinkerTools microworlds or real-
world situations, and they evaluated the limitations of their model and their investigation. The prior ThinkerTools students did not evaluate their models with the specific criteria of accuracy, plausibility, utility, and consistency, nor did they engage in scientific debates to determine the best explanatory law. Rather, they engaged in a class discussion in which they tried to reach a consensus about the most accurate law.

**Research Projects**

The final METT module (as well as the final module in the prior version of the curriculum) did not follow an identical format to the other modules. In this module, students conducted their own research projects. They divided into groups of two to four students and picked one of three different possible topics to investigate: mass and the effect of an impulse, gravity, or gas/fluid resistance. (Students in the prior version of the curriculum chose from among eight topics.) Once they had conducted their research, students in both versions of the curriculum presented their results in oral presentations to the rest of the class. Students in the METT classes then evaluated each other’s presentations with the modeling criteria as well as the general criteria for judging research (being inventive, reasoning carefully, and so on) used in the prior curriculum.

All students in both versions of the curricula were then given post-tests including a physics test, an inquiry test, and a scientific beliefs test. METT students also took a modeling knowledge test.

**Experimental Design**

To reiterate, this study addresses three primary questions. What do students learn about the nature of scientific models and the process of modeling from curriculum and instruction that
focuses on model creation, evaluation, and reflection? What effect does model-centered
curriculum and instruction have on students’ inquiry skills and conceptual physics knowledge?
And, how is the development of modeling knowledge related to the development of physics
knowledge and inquiry skills? We note that this study also looked at the effect of the METT
curriculum on students' epistemological beliefs, and these results are reported elsewhere
(Schwarz, 1998).

Our primary results that bear on these research questions are gain scores between pre-
instructional and post-instructional assessments of students’ understanding of modeling, inquiry,
and conceptual physics, as well as correlations among these scores. Copies of these assessments
can be found in the first author’s doctoral dissertation (Schwarz, 1998). In addition to calculating
gain scores, we carried out post-instructional interviews with twenty-two students on the nature
of models. We conducted the interviews in order to determine the depth of students’ modeling
knowledge and the retention of modeling knowledge several months after the end of the
curriculum.

The methodology in this research incorporates aspects of a design experiment (Brown,
1992; Collins, 1992). The Model-Enhanced ThinkerTools curriculum is a significant revision of
the prior ThinkerTools curriculum. However, because both versions of the ThinkerTools
instructional trials ran for identical lengths of time, with one of the same teachers in the same
school, we were able to compare outcome scores on our inquiry and physics assessments for the
two versions of the curriculum.
Framework For Modeling Knowledge

Before describing the methods used in this study, we first describe our framework for developing and assessing modeling knowledge. The instructional intervention and assessments were designed to address four types of modeling knowledge summarized in Table 1.

We decided on these four dimensions (nature, process, evaluation, and utility) by reviewing the literature (Grosslight et al., 1991; Carey & Smith, 1993) and then analytically determining key dimensions of modeling knowledge. These dimensions are not a definitive way of categorizing modeling knowledge, only a useful framework for creating the assessment and interview questions. As Table 1 indicates, the first type of modeling knowledge is knowledge about the nature of models. This dimension encompasses understanding what models are, what models represent, and that there can be different kinds of models for the same object or phenomenon. Our second dimension, the nature of modeling, encompasses knowledge about the process of modeling. In other words, understanding that modeling involves a process of embodying key aspects of theory and evidence into a model, testing that model and then revising it. The third dimension of modeling knowledge involves evaluation of models, including the idea that there are ways to evaluate models in order to determine whether one model is better than another and that there are criteria, like accuracy and plausibility that can be used in this evaluation. Finally, the fourth dimension involves understanding the utility of models and
understanding how models can be useful for developing and testing theories and for visualizing and explaining phenomena.

METHOD

Setting and Participants

Two middle school science teachers and the first author implemented the Model-Enhanced version of the ThinkerTools Inquiry Curriculum in eight seventh-grade classes in an urban San Francisco Bay Area public middle school during the 1996-97 academic year. The population of the school was ethnically, economically, and academically diverse. For example, during the 1996-97 academic year, approximately 44% of the school’s students were Black, 31% were White, 13% were Asian, 11% were Hispanic, and 1% of the population was composed of other groups. Additionally, thirty-four percent of students qualified for free or reduced lunch, and 20% came from families who received Aid for Dependent Children. Students in this study displayed a wide range of scores on the Individual Test of Academic Skills (ITAS) ranging from the 7th to 99th percentile. (ITAS is a nationally normed test that was given to California school students during the 1996-97 academic year.) The ITAS distribution for these classes had a median percentile score of 66. This median was higher than the median percentile score of 60 on the Comprehensive Test of Basic Skills (CTBS) from the trial of the previous ThinkerTools curriculum.

The first author co-taught one class of the curriculum with teacher A and was a teachers’ aide in one of teacher B’s classes. The two teachers, A and B, used the curriculum without the first author’s presence in their remaining six classes. Teacher A used all four modules for approximately ten and a half weeks in all four of her seventh grade classes and teacher B used
module 1 (the effect of friction on one-dimensional motion), part of module 3 (two-dimensional motion), and module 4 (student projects) for approximately seven and a half weeks in his four seventh grade classes. Teacher A was a twenty-five year veteran teacher who had taught the prior version of ThinkerTools for three consecutive years. Teacher B was in his third year of teaching middle school science, had not taught any ThinkerTools curricula prior to this study, and preferred having the first author take a more minor presence in his classroom. Because teacher B was a less-experienced teacher who chose to use a smaller portion of the curriculum, and because he had introduced some of the curriculum’s modeling ideas to his classes prior to the beginning of the study, we chose to analyze only two of his four classes. We felt that his outcomes were less representative of the overall impact of this type of approach. Nonetheless, data from his two classes may provide additional insights into the nature and generalizability of the outcomes, so we thought it important to report data from some of his classes. We note that the presentation of the results will combine teacher A and B data unless there are significant differences between the teachers, in which case we will present those differences.

**Instruments and Analyses**

While conducting this study, we used several different assessment instruments and data sources to address our research questions. These include paper and pencil pre-tests and post-tests, student interviews, student work, classroom videotape, and the first author’s reflective journal. For the purposes of addressing our research questions, in this paper we focus on reporting results from our written pre-post-assessments and from the interviews with students after the curriculum. The three written paper-and-pencil pre-post-tests include a modeling assessment, an inquiry test, and a conceptual physics assessment. As previously mentioned, we
also administered an epistemological beliefs questionnaire whose results are reported elsewhere (Schwarz, 1998). Each written assessment was administered during one class period and did not affect students’ grades in the class. In addition, the first author conducted nature-of-models interviews with twenty-two students from two classes (one teacher A and one teacher B class) two and a half months after the curriculum and instruction ended.

**Written Assessments**

The Modeling Assessment included questions about the nature, evaluation and purpose of models and utilized a variety of question formats. These formats include a sorting task (circling all the items that are models), enhanced multiple-choice questions (e.g. “What is the best definition of a model and why?”), and enhanced true/false questions (e.g. “Could a scientist create an incorrect model and why?”). For a more complete list of items from that assessment, see Appendix A.

The written Scientific Inquiry Assessment was designed for instructional trials of the previous version of the ThinkerTools curriculum (White & Frederiksen, 1998). In this assessment, students are asked to conduct an investigation that follows the inquiry cycle shown in Figure 4. The assessment begins with the research question, “What is the effect of varying the weight of an object on what sliding friction does to its motion?” In this assessment, students are asked to develop and justify alternative hypotheses about possible answers to the question, design an experiment to test their hypotheses, create made-up data that they might get if they ran the experiment, analyze those data, draw conclusions, and then relate the conclusions back to their hypotheses.
The written Applied Physics test, a subset of which was also used in the trial of the previous ThinkerTools Inquiry curriculum (White & Frederiksen, 1998), included sixteen enhanced multiple choice items that assess physics concepts ranging from knowledge about one-dimensional motion with and without friction, to two-dimensional motion and gravity. For example, in one sample item, students were asked the question in Figure 7 below.

---

Insert figure 7 about here

---

**Analyses of Written Assessments**

After calculating students’ overall scores on each assessment, we used paired t-tests to determine whether the statistical means of the assessments were significantly different before and after the curriculum. Differences in the success rate on individual items before and after the curriculum were determined using McNemar chi-square tests of significance. We also used two-sample t-tests and McNemar chi-square tests to compare results from the Model-Enhanced ThinkerTools to results from the prior ThinkerTools Inquiry curriculum. Furthermore, we analyzed data from all of the written assessments using analyses of variance (ANOVA) that included five between-subjects factors. These factors included: teacher (A or B), academic achievement level (low [ITAS score < 60], high [ITAS score > 60]), prior ThinkerTools experience from the sixth grade (yes, no), gender (female, male), and the first author’s presence in the classroom (yes, no). We also included the time of the test (pre and post) as a repeated measures factor.
This paper also presents correlations among the written assessment scores. These correlations indicate how students’ performances on the modeling, inquiry, and physics assessments are related to one-another, and whether modeling knowledge and its acquisition is related to the development of physics knowledge and inquiry expertise.

**Modeling Interviews**

In order to determine the depth of students’ modeling knowledge and the amount of knowledge students retained, the first author conducted modeling interviews two and a half months after the curriculum and instruction ended. Analysis of the Modeling Interviews determined the frequency, quality, type, and distribution of students’ responses to the interview questions as well as the factors that contributed to the distribution of responses.

The Modeling Interview included questions relating to the four types of modeling knowledge shown in Table 1 and lasted between thirty to fifty minutes. The interview included contextualized questions about models and modeling related to students’ final projects (“Did you get a chance to try out the different rules for your research findings in the modeling step of the inquiry cycle? Why should a student do this?”), decontextualized questions about the nature of models and modeling (“In general, is any model just as good as another?” “Do scientists ever change or revise their models?”), and two activities, (“Here are two examples of scientific models of gravity. How would you decide which is the best model?” and “Suppose that you wanted to find out how long it takes for a student to get between classes at your school. Using the inquiry cycle, describe how you might investigate this question.”)

The first author conducted the interviews by adhering to the interview question format and question phrasings, although in a flexible manner. For example, questions were occasionally
phrased or ordered differently so that they could more naturally fit into the conversation.

Further, the interview format had the flavor of a clinical interview in which the first author occasionally prompted students by asking them to explain their responses or by reminding them of additional evidence. Additional details regarding the analysis and coding of the Modeling Interview can be provided to the reader upon request (Schwarz, 1998).

Twelve of the interview subjects were students in one of teacher A’s classes, and ten were students in one of teacher B’s classes. Further, fourteen students or 64% of the interview sample were female, while the remaining students were male; fifteen students or 68% of the interview sample had ITAS scores higher than 60, the remaining seven students had ITAS scores lower than 60. We interviewed all students from those two classes who volunteered and received parental permission to spend the time away from class.

**Interview Analysis**

Interviews were transcribed from audiotape. From these transcriptions, the first author summarized students’ statements into a summary sheet. This summarization was a data reduction technique used to condense and capture an abbreviated version of the students’ responses while maintaining as much fidelity to the students’ responses as possible (see Schwarz, 1998 for more detail).

Once the interview transcripts were summarized, we created a coding scheme so that we could characterize and then aggregate the students’ responses to each interview question. First, we created comment codes to characterize each distinct idea or statement a student might give in response to the interview questions. We also categorized each comment code into one of three levels (strong, moderate, or weak – also called level ratings) based on judgments about the overall
sophistication of the responses that fit that comment code compared to responses that an expert
might provide (such as those from the second author and her colleagues) as well as to the level
ratings given to such responses by other researchers (Carey & Smith, 1993; Grosslight et al,
1991). Samples of the comment codes and level-ratings are shown in Tables 3-7.

We then assigned comment codes and level-ratings to students’ interview responses. As
students' responses were usually lengthy, we often assigned multiple comment codes to a given
response. Subsequently, each student was given an overall level rating (strong, moderate, weak)
for each interview question to characterize his or her overall response. These level ratings were
based on either an average of their level ratings for their responses to that question or their
highest level rating for their response to that question. (See individual tables 3-7 for indications as
to how the overall level ratings were determined.)

We then tested the reliability of coding the interview summaries compared to the
reliability of coding full transcripts for both the content categories and level ratings. We found
that the summaries provide a reliable basis for rating students’ responses, giving an average of
90% agreement with the coding of the full transcripts. As a result, we analyzed the ratings of
interview summaries instead of the full transcripts.

In our analyses of interview results, we determined the frequency of interview responses
as well as the significance of factors such as teacher, academic achievement level, prior
ThinkerTools experience, and gender using Fisher Exact Tests. We also calculated the correlation
between students’ overall interview rating (as determined by adding up their overall level rating
for each of the eleven interview items) and their scores on the modeling post-test in order to
determine the consistency of responses between these two assessments.
RESULTS – MODELING KNOWLEDGE

In this section of the paper, we turn to addressing our primary research question, what did students learn about the nature of scientific models and the process of modeling from the METT curriculum and instruction, which focused on model creation, evaluation, and reflection? In order to address this question, we determined the effect of METT on modeling knowledge by comparing pre-test with post-test scores from the Modeling Assessment and by analyzing the type and quality of modeling knowledge students retained after the curriculum, as revealed in their answers during the Modeling Interview. Modeling knowledge was not assessed with either the Modeling Assessment or the Modeling Interviews in evaluations of prior versions of the curriculum and so could not be compared.

General Modeling Knowledge

We analyzed overall scores on the Modeling Assessment for six classes (four teacher A classes and two teacher B classes) on the eighteen multiple choice items of the assessment that are common to both the pre-test and post-test. A summary of results from this analysis is given in Table 2.

Insert table 2 about here

Overall, there was a significant improvement in students' modeling knowledge. Students had an average pre-test mean of 64% correct (SD=13) and an average post-test mean of 71% correct (SD=14) ($t_{103} = 5.95, p < .001, \Delta = .50$). In carrying out a repeated measures analysis of
variance, with time of testing (pre-test and post-test) as the repeated measures factor, we found that there was no significant effect of gender or of the experimenter’s presence in the classroom. However, we found that there were significant effects of students’ academic achievement level and of teacher. With respect to academic achievement (as determined by scores on the ITAS assessment), students with lower academic achievement showed a mean gain of 1% compared to a mean gain of 8% for students with higher academic achievement ($F(1, 104) = 8.29, p = .005$). We also found that teacher A’s students had a pre-test mean of 61% and a post-test mean of 68%, which is significantly different from teacher B’s students who had a pre-test mean of 67% and a post-test mean of 69% ($F(1, 104) = 5.42, p = .02$). Teacher B’s students began with a pre-test mean roughly equivalent to the post-test mean for students in teacher A classes.

This result for teacher B’s students may be attributed to his heavy emphasis on the role of models before he began the METT curriculum. Upon later investigation, we determined that teacher B decided to incorporate an emphasis on understanding modeling in all his science units after becoming aware of its importance in talking with the authors prior to the beginning of the school year. As we were unaware of this decision, we did not document this prior modeling instruction. However, we do know that he taught students about modeling within the context of his astronomy unit which lasted approximately ten weeks and was taught prior to METT. Astronomy models at this grade are usually presented as concrete physical models that explain concepts related to the solar system. In addition, teacher B engaged students in a shorter version of the METT curriculum that lasted only 7.5 weeks rather than 10.5 weeks, thus giving students less opportunity to increase their understanding of modeling further. These instructional differences between the two teachers need to be born in mind as we analyze students' responses
to items in the Modeling Assessment and the Modeling Interview questions in the following sections.

Analysis of Modeling Knowledge Dimensions

Dimension I: Nature of Models

Analysis of Modeling Assessment

What did students understand of the nature of models? Did students understand what a model was, that there can be multiple models of the same object or phenomenon, and that models are constructed? We will begin with an analysis of individual items from this dimension on the Modeling Assessment and subsequently report results from the modeling interviews. Our hypothesis is that students should improve their understanding of models due to the emphasis in METT on reflecting on the nature of models and interacting with multiple forms of models, particularly the predictive and explanatory rules that governed the ThinkerTools microworlds. We also hypothesize that if students develop a deep understanding, they should continue to show evidence of understanding about the nature of models several months after the end of the curriculum and instruction as they undergo the Modeling Interview.

For the seven nature-of-models questions within the Modeling Assessment, students scored an average of 67% in the pre-test and 73% in the post-test ($t_{103} = 5.16, p < .001, \square = .48$). Most of this change was due to teacher A’s students who improved their scores an average of 9% ($t_{71} = 5.69, p < .001, \square = .67$) compared to teacher B’s students who improved by 3% ($t_{31} = 1.0, p = .31, \square = .15$), partly because teacher B’s students started the curriculum with high scores on
more than half the items for this dimension – they were able to correctly identify many of the concrete models both before and after METT.

For example, the first question of the written Modeling Assessment asks students to identify which of eighteen items they think are models. The items ranged from non-models, such as a pencil or bicycle, to concrete visual models, such as a globe or a drawing, to more abstract models, like scientific rules and theories. There were significant changes to students’ categorization of all the items that were models, except for items that students already believed were models (such as a toy car or a person who displays clothes) on the pre-test. There were no significant changes for other items, such as the pencil, bicycle or tree, which the majority of students agreed in the pre-test were not models. The most dramatic results were changes for teacher A’s students’ categorizations of the most abstract models, such as a causal rule, an equation, and a scientific theory. For example, in teacher A’s classes only 14% of students thought a causal rule was a model in the pre-test, whereas 48% of them believed it was a model in the post-test ($\chi^2 (1, n=71) = 22.15, p < .001$). Similarly, in teacher A’s classes, only 21% of students in the pre-test thought an equation like Newton’s second law was a model compared to 55% of students in the post-test ($\chi^2 (1, n=71) = 22.15, p < .001$). Finally, only 21% of students in teacher A’s classes thought a scientific theory was a model in the pre-test, compared to 52% of students in the post-test ($\chi^2 (1, n=71) = 17.29, p < .001$).

These changes were not significant for teacher B students. Thirty-eight percent of teacher B’s students considered a causal rule to be a model in the pre-test, compared to 40% in the post-test ($\chi^2 (1, n=32) = .07, p = .8$). Similarly, 22% of teacher B’s students thought of an equation as
a model in the pre-test compared to 34% in the post-test ($\chi^2 (1, n=32) = 1.3, p = .25$). Finally, 35% of students thought of a scientific theory as a model in the pre-test compared to 25% of teacher B’s students in the post-test ($\chi^2 (1, n=32) = 1.8, p = .18$). The lack of a significant gain for Teacher B students may be due to the earlier exposure to concrete astronomical models in Teacher B’s class, making it more difficult for students to understand models as more abstract representations. Alternatively, it may be due to the shortened version of the METT curriculum that Teacher B students received.

Students showed improvement on several additional nature-of-models assessment items including the question below:

Analysis of the entire sample (four teacher A classes and two teacher B classes) indicated that students significantly improved their understanding of the definition of a model. Thirty-six percent of students chose either ‘B’ or ‘C’ in the pre-test compared to 53% in the post-test ($\chi^2 (1, n=91) = 6.43, p = .01$). This difference was mainly due to an increase in students choosing option 'B': fifteen percent of students chose ‘B’ in the pre-test compared to 30% in the post-test, whereas 21% chose ‘C’ in the pre-test compared to 22% in the post-test. It is interesting to note that even in the post-test, 16% of students still picked the response, they “really didn’t know.” This shows that, while many students improved their answer to this question, this improvement
was not universal. Such a result is consistent with other research showing that modeling knowledge is difficult to develop (Carey & Smith, 1993).

**Analysis of Interviews**

In the Modeling Interview, students were asked three groups of questions about the nature of models. The first group of questions for this dimension of understanding about models related to students’ definition of a model, the second to multiple models, and the third to the constructed nature of models. Our goal in this analysis was to determine students’ understanding of the nature of models through their own talk about models several months after the curriculum ended.

**Students’ definitions of a model.** We analyzed students' responses to several interview questions to determine students’ definitions of a model. In the beginning of the interview, when students talked about their research project, they were asked, “In this context, what do we mean by a model?” Later in the interview, after the students had completed an activity in which they evaluated two written models of gravity, they were asked, “Now that we’ve talked about it a little bit, can you tell me what you think scientific models are in general?”

We hypothesized that students would improve their conception of what a model is so that it would include entities that are more abstract than toy cars, such as the predictive and explanatory rules they worked with most closely during METT. Prior research cautions us that this is a difficult goal to achieve (Carey & Smith, 1993), and so creating a pedagogical approach that succeeds may require a process of iterative refinement.
Results from our analysis of students’ responses are given in Table 3 below. We note that the second author blindly coded ten randomly chosen interview summaries and obtained an 80% agreement with the first author on the overall ratings.

This analysis of the interview responses indicates that, several months after METT ended, students showed a great deal of sophistication in their overall definitions of a model, replicating the findings from the Modeling Assessment. We found that 73% of students gave a strong definition of a model, and 14% gave a moderately strong definition. One hundred percent of students in teacher B’s classroom had a strong overall rating, compared to 67% of teacher A students (Fisher Exact, $p = .07$, $\square = -.43$). Perhaps this is another effect of teacher B’s emphasis on models in his classroom throughout the entire school year.

The most common of the model definitions students gave was the strong response that a model provides a predictive, explanatory rule (see Table 3). Additionally, several students mentioned that a model is not necessarily a physical object, but could be a mental object or made of words. Some students also included less sophisticated concrete definitions of a model such as a smaller copy or version of an object. The frequency of strong or sophisticated responses is particularly impressive given that the interviews were conducted two and a half months after the end of the curriculum.

To give the reader a sense for how the authors coded the student summary responses, we present two short excerpts of student responses following by their coding categories.
A Model-Centered Approach

A portion of PL’s response rated as strong: “[A scientific model] can be a theory or rule about what you think happens in real life, or it can be a representation of something. Any representation of a real thing like a car model, or a theory.” Response coded as A3, A1, A2.

A portion of IT’s response rated as strong: “A model is a small design or remake of something. Something made over, but smaller. A globe would be an example. ... They don’t always show you anything, but they say maybe that um, ‘Jupiter’s red spot is three earth’s and it’ll keep on going around and around for a long time.’” Response coded as B1, C2, A1, A6.

Students’ understanding of multiple models. We analyzed several questions to determine student’s knowledge about multiple models, which is another subcategory of dimension I – students’ understanding of the nature of models. Specifically, students were asked, “Could someone else who did that same research project come up with a different model? Why or why not? Could a different research group draw different conclusions and create a different model even if they had the same experimental data?” We hypothesized that students would be able to understand that multiple models can exist as they had worked with multiple models for each concept or physical principle addressed by the METT curriculum.

Our analysis of the interview responses indicates that students showed a fair amount of sophistication about multiple models. We found that thirteen of twenty-two students (59%) had an average response that was strong, capturing the sense that different people may have different interpretive frameworks for the same data, while five (23%) had a moderate level response which captured the sense that different models arise because of aspects such as variability and error.
While a few students claimed that multiple models could not exist, most students thought that multiple models could exist because different people conducted their experiments differently or because different people have different opinions or ideas about the phenomena. For example, one student gave this thoughtful response:

A portion of SB’s strong response: “It’s possible [that somebody else who did our same research project could have come up with a different model.] I mean, ... it depends on like how they dropped ... [the marble] and what they saw on the videotape. ‘Cause it also depends like how many little you know [dotprints] they did. ‘Cause you know if they ... did 15 right, it would have looked different. ... It’s possible [that somebody else could have come up with a different model even with the same data] ‘cause it’s all like in the minds eye what they see, what they think they see. So they could have like thought ‘stayed at a constant speed.’ ...[and it] would have been totally different.”

Nature of models - constructed nature of models. We also explored students’ understanding of the constructed nature of models, another subcategory of the first dimension of model understanding. Specifically, students were asked, “Do you think that scientific models tell you what actually happens in a physical situation, the reality of the situation, or do they give us a general idea or estimate of what happens?”

For this sub-dimension, we hypothesized that students would understand that models are constructed and do not necessarily represent an absolute reality. We predicted this outcome because, in the METT curriculum, students are asked to choose one of several conceptual models made available by the software. In evaluating that model, students should notice that their chosen model does not exactly predict their real-world data.
Our analysis of the interview responses indicates that students showed thoughtfulness and understanding about the constructed nature of models. Only three students (14%) gave weak responses stating that they needed to personally see evidence in order to believe something was true. Seven students (32%) gave moderate and more relativistic responses indicating that models are estimates because they are representations of the real world. The majority of students (55%) gave a fairly sophisticated response indicating that scientific models can be very accurate, but are not exact copies of reality (they are human constructs), and that models for which we have the most amount of evidence are more likely to accurately represent the world compared to those that do not. For example, one student gave this strong response:

A portion of EH’s response: “[Models] don’t tell ... exactly what will happen, but they’ll tell us something that might wind up happening. Like if you mix two certain substances together that it might not blow up, but it might. ... We know that there are such things as atoms. Then [the model] gives us an idea of how some, it doesn’t actually show us how they look or anything. It’s just an ... idea of how an atom would look.”

We also note that the teacher was again a significant factor in students’ interview responses regarding the constructed nature of models. Thirty-three percent of students in teacher A’s class received a strong score compared to 80% of students in teacher B’s class (Fisher Exact, $p = .04, \bar{d} = -.47$). This results is consistent with our informal classroom observations which suggest that more of teacher A’s students thought of science in somewhat relativistic terms – that models are constructed without necessarily having any similarities to the real world. This relativistic orientation for teacher A’s students will be further discussed in reporting results of the third dimension of modeling knowledge – the evaluation of models.
Dimension II: Nature of Modeling

What did students understand of the overall modeling process? Did students understand that models are created and often revised with new insights or new evidence?

Our hypothesis was that the METT curriculum, in which students evaluate alternative models, would enable students to understand how models are created and to appreciate that a model might need to be revised if its behavior does not match data from real-world experiments or if the model is inconsistent with other models or data. Analysis of data from appropriate items on the written Modeling Assessment and the Modeling Interview enable us to determine whether this was indeed the case or whether our instructional approach needs further refinement.

Analysis of Modeling Assessment

Analysis of three questions related to the process of modeling in the Modeling Assessment indicated that students scored an average of 52% in the pre-test and 56% in the post-test ($t_{103} = .62, p = .54$). This suggests that students had a moderate understanding of this modeling dimension prior to beginning the METT curriculum and that there was no significant improvement in students’ understanding in this area due to the curriculum and instruction. Another possibility is that these assessment items are not sensitive or valid enough to measure the improvement that may have occurred. An example of one of the questions about the nature of modeling is the following, “If a scientist wanted to create a scientific model of an atom in order to predict how that atom will interact with other atoms, what parts of the atom would a scientist include in the model?” Forty-seven percent of students chose the best response, “only parts useful for predicting how it will interact with other atoms” in the pre-test, compared to 46% in the post-test. The alternative responses included “every single part of the atom” for which 31%
chose this in the pre-test compared to 42% in the post-test and “only the main parts of the atom” for which 22% chose this in the pre-test compared to 12% in the post-test. We conjecture that the increased response rate for “Every single part of the atom” in the post-test is a consequence of participating in METT which had students simulate the real world by adding increasingly complex features. In this sense, students learned exactly what the curriculum demonstrated – that in order for the simulation to be as accurate as possible, it needed to include many real-world features such as friction and gravity.

**Analysis of Modeling Interviews**

Our interview data reveal students’ overall understanding of the process of modeling in response to a question relating to model revision, which is one subcategory of the process-of-modeling dimension. Near the beginning of the interview, after students discussed their research project, they were asked, “Do scientists change or revise their models? If so, why? When would they not change their models? Might scientists change their models even without more experimental data?”

We hypothesized that METT students would learn that scientists revise their models in order to address new insights or to accommodate data just as METT students did in revising their conceptual models within the curriculum. Our computer modeling tool and curriculum was designed to help students re-think the implications of their conceptual models, derived from their real world experiments, and possibly revise their models.

Students’ responses were assigned a score of strong, moderate, and weak based on the most sophisticated response they gave to the question. A strong response indicated that revision may occur by rethinking one’s data and their implications as well as the purpose of the model. A
moderate response captured the sense that model revision occurs when there is new information or mistakes are discovered. A weak response was often a vague one in which the student stated that model revision may not occur or would occur because the model was wrong. Results of our analyses are given in Table 4.

Our analysis of the interview responses indicates that students showed a moderate level of sophistication about revising models. We found that seven of twenty-two students (32%) had a response that was classified as strong while eleven (50%) had a moderate level response, and four (18%) had a weak level response. Students’ degree of sophistication about model revision may have been limited because the curriculum emphasized model creation and reflection more than model revision. While students revised their hypotheses (their initial model) to create their final model during the curriculum, the curriculum and instruction did not highlight the idea that students were building on a simple model to create more complex and sophisticated ones. Further, students might have had a better understanding of model revision had they revised a particular model over several iterations.

By far the most frequent statement given was the moderate response that scientists would change their model to make it more accurate if they found out new information or evidence. We conjecture this response was the most frequent because this was the most common experience of students in revising their own models during the curriculum. While the curriculum provided many opportunities to re-think a model and discuss its implications with other...
students, most of the students focused on trying to revise their initial hypotheses on the basis of new information they gathered from their real world experimentation. The following interview excerpt illustrates a strong response to these interview questions.

A portion of DC’s strong response: “Yeah, [scientists change or revise their models] because they might do the same thing [as me] and go home, or whatever and be thinking about what they did in lab, or something. And be like, “yeah, that’s a big mistake” or something. And then want to come back to the lab and change their idea because they thought that maybe they were wrong and it actually was slower or faster or something.”

Dimension III: Evaluation of Models

What did students’ understand about the evaluation of models? Did they understand that models are evaluated with specific criteria that help determine their merits? Did they understand that even though models are constructed, some are better at approximating the behavior of the world than others?

We hypothesized that students in METT would understand that models can be evaluated with criteria such as accuracy and plausibility, because they would be evaluating their own models with such criteria throughout the curriculum. Also, we hypothesized that students would come to realize that some models more accurately approximate the behavior of the world or are more plausible than others, because these issues manifested themselves when students evaluated each other's models and discussed them in the class debates.

Analysis of Modeling Assessment

Students showed no improvement in this dimension of model understanding on the Modeling Assessment. For the three questions related to the evaluation of models, students
scored an average of 52% on the pre-test and the post-test. For example, most students agreed in
the pre-test (87%) and post-test (94%) that “When a scientist evaluates a scientific model, she
looks for certain qualities such as how accurate and reasonable the model is.” In another item,
students were asked whether they agreed or disagreed with the statement “Since scientists
disagree about why dinosaurs became extinct, it’s clear that no one understands exactly how it
happened. Therefore, any scientific model or theory of how it happened is just as good as any
other.” In the pre-test, 40% of all the students disagreed with this statement compared to 34% in
the post-test. While, there was change in the desired direction in students' responses to both
these items, neither change was statistically significant. In the case of the first item, this may be
at least partly due to a ceiling effect. We need to create a larger set of better items to detect any
changes that occur in students' understanding of model evaluation.

**Analysis of Modeling Interview**

The Modeling Interview gives a clearer indication of what students learned with respect
to model evaluation. In this subsection, we report results from two questions from the interview
that related to model evaluation – the first related to how criteria are used to evaluate models, and
the second is about the value of alternative models.

To determine students’ understanding of evaluation criteria, students were asked, at the
beginning of the Modeling Interview, in the context of discussing their research project, how they
evaluated their research findings and what criteria they used. At the end of the interview,
students also evaluated the model that they created in their thought experiment in which they
conducted inquiry related to the question of how long it takes for students to get between classes
at school. Responses to both questions are included in our analysis presented in Tables 5 and 6.
We categorized students’ responses by looking for evidence of specific evaluation criteria or other means of evaluating models. A strong response was one in which students evaluated their models using particular criteria or compared it to other models. A more vague response, which mentioned notions of correctness or clarity, was given a moderate rating. A weak response was one in which the student expressed vague notions of evaluating or checking work. Students received a code for both model evaluation tasks, and those codes were then averaged.

Our analysis of students’ responses, presented in Table 5, indicates that students generally had a strong understanding and use of model evaluation criteria. Eleven of the twenty-two students (50%) who participated in the Modeling Interview mentioned and used specific criteria to evaluate their models, obtaining a strong overall rating, while seven students (32%) evaluated their model with respect to vague notions of correctness or clarity, receiving a moderate rating, and four other students (18%) only vaguely checked their work, receiving a weak overall rating.

We found that students’ responses frequently mentioned using the model evaluation criteria introduced in the curriculum, including accuracy, plausible mechanism, consistency, and utility. Students also frequently mentioned checking their data or model to see if it made sense, and checking their model to see whether it was “okay.” While students’ performance on these questions may have been due to their recalling the evaluation criteria, these results nonetheless...
suggest that many students understood and used criteria in evaluating their models and are able to recall them several months after the curriculum.

The following interview excerpts illustrate some examples of students’ responses to this interview question that were rated as strong.

A portion of PL’s strong response: “I looked [to see] if my data seemed like plausible, and looked to see if my model could be used to predict anything really accurately.”
Response rated as A1.

A portion of ER’s strong response: “We [evaluated our model by] ... looking at it, and we kind of just thought about it, and we looked at the computer models, and saw if like, which one matched and stuff. ... And we looked at the Newtonian model to see if that was the same as our thought. And, we just kind of looked and see, did it look right to us, and did we think all our research was accurate.” Response rated as A2, A3, A1, C1.

Our second research question for this dimension of expertise was, what did students understand about the relative value of alternative models? To determine this, we asked students the question, “In general, is any model as good as another?” roughly a third of the way through the interview.

Students’ responses were again categorized with respect to their sophistication. A strong response indicated that some models are better able to predict and explain phenomena than others. A student claiming that all models have equal value, particularly if evidence cannot distinguish among them, received a moderate rating. A poor response was an extreme relativistic position that stated that all models have equal value, and there is no method for judging the ‘goodness’ of one model from any other. Results of our analyses are presented in Table 6.
Our analysis of students’ responses indicates that, while some students indicated a sophisticated understanding, many students expressed a naive view of the value of alternative models. We found that eight of twenty-two students (36%) gave a strong response stating that some models are better than others because of their validity or accuracy, while some models are odd or implausible. Two students (9%) gave a moderate-level response that models have equal value when there is no way [for the student] to know which one is right. Ten other students (45%) gave a weak response, holding the relativistic notion that all models have equal value. Students in this category gave a variety of reasons to justify this response, including that people have worked hard on creating their models, people have different opinions, and there ultimately is no right answer. The remaining two students had responses that were unclear.

The following interview excerpts illustrate several students’ responses to these interview questions.

A portion of TH’s weak response: “Yes [any model is as good as another] because models aren’t answers. They’re just possibilities. ... They’re not facts ... They’re thoughts. ... [It has to be] a proven scientific fact. ...it’s like whenever there is a fact, it’s most likely ... in a book, or… (laugh) it is! … Like the school gets all facts from a book. ... They’re just like here. And we have to know them to get to the next grade.” Response rated C7.
A portion of ER’s strong response: “I don’t think [any model is just as good as another].
Because some people ... don’t really do their research that well. ... If there is one model
that the research is accurate, and the people were pretty smart about their hypotheses
and their conclusions and everything, and then the other people were just kind of sloppy
about it, then that model isn’t as good as the other one.” Response rated A1.

It appears from the analysis of these interview questions about the evaluation of models
that some students’ behavior in evaluating their models was inconsistent with their views about
the value of models. That is, most students understood and used evaluation criteria, but many
stated that all models have equal value. It is possible that students who gave relativistic
responses interpreted the interview question as asking whether all models have value, as opposed
to accuracy and plausibility, and were unwilling to discuss the relative value of one model over
another – a social belief that was nurtured in other aspects of the school environment. It is also
possible that because the different computer modeling rules in the last two modules were difficult
to evaluate based on the criteria, we inadvertently moved students towards thinking of all
scientific models as being plausible and therefore accurate – in which case, we need to revise our
instructional materials.

Dimension IV: Purpose or Utility of Models

Finally, we assessed students’ understanding of the purpose of models in order to
ascertain whether students understood the purpose and utility of modeling. We hypothesized
that students would develop a strong understanding of this aspect of modeling expertise, because
METT engages students in a variety of discussions about how models are used in science, and
gives students many opportunities to use their own and competing models to predict and explain various force and motion phenomena during the curriculum.

**Analysis of Modeling Assessment**

Analysis of questions on the Modeling Assessment suggests that students had a good understanding of the purpose of models by the end of the METT curriculum. In one question about the purpose of models in the Modeling Assessment, for instance, students were asked, “From a scientific point of view, which is the best use of a model?” On the pre-test, 48% of students chose the most sophisticated answer, D, “to develop and test ideas” compared to 61% in the post-test ($\chi^2 (1, n=82) = 5.76, p = .016$). Other possible responses to this question included, “to be a toy,” “to copy an object or process,” or “to help someone construct and object.”

There was, however, a difference for Teacher A versus Teacher B students regarding the change from the pre-test to pos-test. Overall, on the five questions related to the purpose of models, students scored an average of 63% in the pre-test and 68% post-test in the post-test ($t_{103} = 1.4, p = .16$). However, Teacher A students began with a mean score of 58% and ended with a mean score of 75% ($t_{71} = 4.4, p < .001$). Whereas, Teacher B students began with a mean of 73% and ended with a post-test score of 53%. This seemingly inexplicable drop for Teacher B students is statistically significant, but it is at least partly due to pre-test ceiling effects for his students on three of five questions.
Analysis of Modeling Interview

In order to determine students’ understanding of the purpose of modeling in the interview, the students were asked three questions: “What do you think computer or scientific models can be useful for? How do you think computer models can be useful for scientists? How do you think computer models can be useful for learning about science?” A student’s responses to all three questions were categorized and then given an overall rating using a three level rubric according to how well she understood the utility of modeling. For example, if part of the student’s response mentioned specific purposes of modeling, such as envisioning, predicting, testing, or teaching, she received a ‘strong’ rating. A more vague response indicating that models may be useful for constructing an object, to get information, or to make life easier was given a moderate rating. A weak response was one in which the student did not think that models could be useful. There was a 100% agreement between the first and second authors when we assigned level ratings to a subset of ten randomly chosen interview summary responses about the purpose of modeling. Table 7 presents the coding scheme and frequencies of responses related to the purpose of modeling.

Our analysis of students’ responses demonstrates that students had a strong understanding of the purpose of scientific and computer models. We found that eighteen students (82%) mentioned high-level purposes such as visualizing one’s own and other people’s ideas,
while three students (14%) understood the purpose of models in a more vague way, and one
student gave no response to this question.

The following interview excerpts illustrate several student responses to these interview
questions.

A portion of PD’s strong response: “[Scientific models can be useful for] explaining how
the world works and everything like that. ...Explaining the way things are in the
world....[A computer model like ThinkerTools could be useful for scientists if] they want
to look at another person’s model, or the real model, they can look at the ThinkerTools
model and see what the differences and similarities [are]. And come up with a good
answer for it.” Response coded as A6, A2.

A portion of PL’s strong response: “[Scientific models can be useful for] sports, anything
mechanical, and making new technology and stuff. ... It’s useful for predicting what you
need to do . ... [Scientific models can be useful for scientists] because then they can
represent all their other stuff, they can do other stuff. They say, you know ... ‘if this is
true, then this must be true too, and I can do this, and this, and this.’ ... We can see our
own models and stuff. ... So we can see stuff that can’t really happen, that we can’t really
see.” Response coded as B1, A4, A1, A2, A7, A1, A5.

Correlations Between Interview Results and Modeling Assessment Post-test

In order to determine the relationship between the Modeling Assessment and the
Modeling Interview results, we correlated the total scores for these two assessments. We found a
moderately strong and significant correlation between the Modeling Interview scores and the
Modeling Assessment post-test scores for the twenty students for whom we had both measures ($r=.50, p = .025$). This significant correlation between the Modeling Interview and the Modeling Assessment post-test may indicate that the written assessment is predictive of the modeling knowledge detected in the interviews, which is a richer assessment format and was obtained several months after the curriculum. This result gives us encouragement for continuing to refine the written assessments about modeling knowledge, because it may be meaningful and useful for classroom teachers and researchers. It also indicates that the written assessment may be useful for supplementing richer assessment measures, such as interviews, that must be conducted with smaller samples of students.

Summary and Discussion of the Modeling Results

In summarizing our findings from both the Modeling Assessment and the Modeling Interview, we found that the pedagogical approach developed for METT promoted students’ understanding of the nature and purpose of models. Specifically, after the METT curriculum and instruction, many students were able to identify abstract models and understand that a model is a representation that predicts and explains. We also found that most students understood that there can be multiple models for the same object or phenomena, that there can be incorrect models, and that models are estimates of the physical world. In terms of students’ understanding of the purpose of models, we found that most students learned that scientific and computer models are useful in a wide variety of ways including for visualization, testing theories, predicting phenomena, helping people understand science, and conducting investigations that are not otherwise possible.
On the other hand, the results also suggest that the METT approach was less successful at promoting understanding about creating, revising, and evaluating models than we had expected. Many students appeared to have limited understanding of how models are created or that scientists can revise models in response to new insights or new data. And, while students clearly learned and used model evaluation criteria, many indicated that they considered all models to be of equal value when asked, "In general, is any model as good as another?" These findings may indicate needed improvement to our curriculum or limitations in our assessments.

Additionally, we found differences in the results for students of teacher A versus teacher B for two of our dimensions of modeling expertise. Teacher A students generally made large improvements in understanding that models can be abstract entities, while teacher B’s students began with a strong understanding of concrete models, but made less improvement in thinking about models as abstract entities. This may have been due to prior exposure to concrete astronomy models in Teacher B's astronomy curriculum. Further, teacher B students made no gains in understanding the purpose of models, perhaps because of their prior exposure to modeling which enabled his students to begin with strong understanding in this area. In contrast, Teacher A's students made significant gains in this area. The comparisons of the pre-test performances of Teacher A’s and Teacher B’s students suggests that there is more than one way to effectively teach about the nature and purpose of modeling, and that the methods teacher B used to teach about modeling within astronomy were effective, at least in regard to these aspects of modeling.

Overall, these results show that students’ understanding about modeling can be improved through model-centered instruction. The improvement of students’ modeling knowledge after
METT may be due to several aspects of the curriculum and instruction. As previously hypothesized, improvement in students’ understanding of the nature and purpose of models may have been a result of class discussion and reflection about these topics. For example, students’ definition of a model probably improved when interacting with and talking about multiple types of models, such as the videotape of modern computer simulation models, the computer model with its various qualitative and semi-quantitative modeling rules, and examples of models that are discussed in the reading passages. Students’ understanding of the purpose of models most likely improved by experiencing situations in which models can be seen to help in a variety of ways, including helping them think about their conceptual models, in running a range of simulations in a computer microworld, and in predicting and explaining phenomena such as the collision of galaxies. We also suspect that students’ improved their general understanding of the nature of models and the utility of modeling by generating models on the computer, in their research books and in their project reports and then using those models to answer various application questions presented at the end of each curricular module.

Why did our data indicate mixed gains for the process of modeling? The abstract nature of the modeling process and the fact that the modeling process (the cycle of creating and revising) was conducted only three or four times during instruction probably played a role. Further, the METT software is only a partial implementation of model-expressive software (as indicated in Figure 1), which does not enable students to build models from scratch. It is possible that using a more fully expressive type of software may have advantages for helping students learn more about what is involved in creating and revising models.
Why did our data indicate mixed gains for the evaluation of models? With respect to difficulties students encountered in evaluating models, we suspect that some of the models students used and created in METT were difficult to evaluate using criteria such as plausibility. As they embodied common misconceptions about force-and-motion phenomena, many of the alternative models that the software made available to students may have seemed plausible. Students were also not used to the type of fine-grained evaluation that they were asked to do in METT and were confused by the lack of being presented with a ‘right’ answer. Additionally, the culture of the class and school contributed to the idea that students must value each other’s ideas and, to some students, this meant each other’s models as well. All of these factors may have moved students towards a more relativistic position in thinking of all models as having equal value.

The finding that students learned about the nature and purpose of models differs from the results of Grosslight et al. (1991). We suspect that our model-centered approach, with an explicit focus on developing students’ understanding of models (reflecting on models, seeing multiple models, enabling students to run computer models with their own laws and observing their implications) over a substantial period of time, effected the change. We also note that the gains made were not uniform across our dimensions of modeling expertise. This finding may indicate that our measures are not uniformly sensitive or that our model-centered approach needs to be revised and that using a uniform developmental description, as in Grosslight et al. (1991), may not accurately represent the evolution in students’ modeling knowledge. We will return to these issues at the end of this paper.
RESULTS – INQUIRY AND PHYSICS KNOWLEDGE

Now that we have seen how a model-centered instructional approach developed students’ understanding of models and modeling, we turn to addressing our second major research question, determining how the curriculum and instruction affected students’ inquiry skills and conceptual physics knowledge. As mentioned previously, it is important to understand the potential benefits and drawbacks of such model-oriented curricula, particularly the effect on students’ inquiry skills and content knowledge, if we are to advocate and improve on this approach. Further, by comparing the outcomes between this version of the curriculum (in which the primary focus was on developing students’ understanding of modeling) and the prior version (in which the primary focus was on developing conceptual models of force and motion as well as inquiry skills), we can determine how the different curricular emphases affected students’ capabilities in these areas.

Inquiry Skills

In this section, we address the question, “How did METT affect the development of students’ inquiry skills, and how do these results compare with those from the prior version of the curriculum?”

The Scientific Inquiry Assessment

As previously mentioned, we used a written inquiry test that has been used in prior ThinkerTools research (see White & Frederiksen, 1998) to assess students' skills in this area. To reiterate, students are given the question “What is the effect of varying the weight of an object on what sliding friction does to its motion?” Students develop alternative hypotheses about the answer to this question, design an experiment to test their hypotheses, make up some data they
might get if they conducted their experiment, analyze those data, draw conclusions about those data, and relate those conclusions back to their hypotheses.

The inquiry test was scored using an analytic method developed in the earlier ThinkerTools research (White & Frederiksen, 1998). This analytic method entailed coding discrete features of the students’ responses with regard to their inquiry skills, such as whether they controlled variables in their experiment appropriately. The accuracy of the physics did not impact the students’ scores. The scoring scheme included five subscores that were weighted and combined to determine a total combined score. These included Hypothesis (15%), Experiment (35%), Results (15%), Conclusion (15%), and Coherence (20%). In the hypothesis section, for example, students’ scores depended on what variables they used, whether they presented an explanation for the hypothesis, and the quality of their explanation. The coherence score was determined by analyzing the extent to which students’ experiments follow from their hypotheses, their made-up data relate to their experimental design, their conclusions follow from their data and relate to their hypotheses.

We hypothesized that the METT approach would improve students’ overall inquiry scores because it is an inquiry-based curriculum in which students are asked to develop hypotheses, design and carry out investigations to test their hypotheses, analyze their data and draw conclusions in the form of a law or model that summarizes their findings. We expected that the METT curriculum would improve students’ performance in developing law-like models for their conclusions, because they had more discussions about the form that scientific models can take and about the criteria by which models should be evaluated than was the case for the prior ThinkerTools curriculum. Finally, we hypothesized that the gain in the coherence scores for
students in METT might be smaller than for the prior ThinkerTools curricular trials as there was less self-assessment in the METT curriculum, and engaging in self-assessment was found to be an important factor in improving students' coherence scores (White & Frederiksen, 1998).

The inquiry tests were all scored blindly so that the scorer (the first author) would not be influenced by the students’ identity or whether the test was taken before or after the curriculum and instruction. A sample of 116 Inquiry Assessments from two teacher A and two teacher B classes were selected for scoring using the criterion that each student selected had to have completed all the pre-tests and post-tests. This resulted in a sample of thirty-eight sets of pre-post inquiry tests for teacher A and twenty sets of pre-post tests for teacher B. The reliability of the scoring was determined through double scoring of a sample of Inquiry Assessments drawn from the earlier ThinkerTools study. The interscorer reliability for the five subscores ranged from .81 (for hypothesis) to .94 (for conclusions), and the reliability of the weighted average of these subscores was .96.

**Overall Results**

Our analysis of students' performance on the Inquiry Assessment indicates that students showed a significant improvement in their inquiry skills after completing METT. We found that students of Teacher A and B, taken together, had a mean of 46% on the pre-test (SD = 24%) and a mean of 58% on the post-test (SD = 22%; \( t_{57} = 4.39, p < .001, \Delta = .5 \)). However, while we found that teacher A students’ gains were significant, those from teacher B’s classes were not (\( t_{19} = -1.466, p = .16 \)), and the difference in their students' gain scores is close to being marginally significant (\( t_{56} = 1.53, p = .13 \)). Results for the Inquiry Assessment are given in table 8.
The lack of overall significant gains for teacher B students may be because teacher B was a less-experienced teacher with no prior experience using the ThinkerTools curriculum and because he only used a portion of the curriculum. We therefore suggest that teacher B’s students’ scores are probably not representative of what would occur when a more experienced teacher, like teacher A, uses the entire METT curriculum.

---

**Insert table 8 about here**

---

How do these results compare to those from the prior ThinkerTools study? Results are reported in Table 8. A comparison of inquiry scores from teacher A's students to the students of a different teacher from the previous ThinkerTools study shows no significant differences in the pre-test means, the post-test means, or the gain scores between the METT curriculum and the earlier ThinkerTools Inquiry curriculum. The students from the previous study came from the classroom of another seventh grade teacher in a different school, so this is not an ideal comparison. However, we note that there was no statistical difference between teacher A’s inquiry post-test scores from the METT trial and those from her students in the 1994 ThinkerTools trial for which post-test but no pre-tests were given to her students.

**Analyses of Subscores of the Scientific Inquiry Test**

In order to determine in greater detail how METT affected students’ inquiry skills and how those results compare to the prior ThinkerTools study, we analyzed the subscores of the Inquiry Assessment. For the reader’s benefit, we have converted all subscores of the inquiry test into percentages that range from 0 to 100. Again, results are reported in table 8.
Hypothesis-generating subscores

METT students in both teacher A and B classes had strong hypothesis-generating skills before the instructional trial and did not show a significant improvement due to the curriculum and instruction. Teacher A students had mean hypothesis gains of 2% compared to 13% gains from the prior ThinkerTools trial ($t_{85} = 1.98, p = .05$), however, there was no statistical difference between teacher A’s post-test subscores in this area for her students in this instructional trial compared to the previous one ($t_{84} = 1.17, p = .24$).

Experimental design subscores

Students in the METT showed a significant improvement in their experimental design skills. Prior to the curriculum, METT students from both teachers A and B combined had a mean score of 40% (SD=22%) on this portion of the test compared to 50% (SD=19%) in the post-test ($t_{57} = 3.68, p = .001, \bar{d} = .42$). Teacher A’s students’ gain in experimental design scores were similar to those for the previous ThinkerTools students, and there were no statistical differences between teacher A’s students post-test subscores in this instructional trial compared to the previous one. In contrast, Teacher B's students showed no significant gain in their experimental design subscores, going from a mean of 35% on the pre-test to a mean of 41% on the post-test ($t_{19} = 1.516, p = .15$), although the difference in the gain scores for Teacher A versus Teacher B students was not significant ($t_{56} = 2.08, p = .31$).

Results subscores

The METT students also showed significant improvement on the results section of the Inquiry Assessment. Students from both teachers A and B combined had a pre-test mean results
score of 29% (SD= 23%) and a post-test mean of 56% (SD=24%; $t_{57} = 7.7, p < .001, \bar{d} = 1.16$). Teacher A’s students’ scores were again similar to those from students in the prior ThinkerTools study. However, gain scores of 32% from teacher A’s students’ were significantly higher than the 17% gain scores from teacher B’s students ($t_{56} = 2.08, p = .04$). This may again be a consequence of Teacher A doing more of the METT curriculum and being a more experienced teacher. The fact that the subscores and the gain in the subscores for her students are almost identical to those for Teacher C’s students is also supportive of this interpretation. Again, we found no statistical differences between teacher A’s post-test scores for her students in this instructional trial compared to the previous one.

**Conclusion subscores**

METT students also showed significant improvements in their skills in the fourth section of the Inquiry Assessment, the conclusion section. Students from both teachers A and B combined began with a mean conclusion score of 40% (SD=35%) and improved to a post-test score of 54% (SD=29%; $t_{57} = 3.03, p = .002, \bar{d} = .41$). Again, teacher A’s students’ scores were similar to those from the previous ThinkerTools students, and her students’ gain scores of 21% were significantly greater than the 2% gain scores from teacher B’s students ($t_{56} = 2.07, p = .043$). Most interesting to note, however, is that teacher A’s post-test scores from METT were significantly higher than those from her trial of the prior ThinkerTools curriculum ($t_{84} = 2.43, p = .02$). This lends some support to our hypothesis that the METT approach may be more effective in enabling students to draw appropriate conclusions and form law-like models from their data. Since this was one of the major goals of the METT curriculum, it is an encouraging
finding.

Coherence subscores

Finally, METT students showed significant improvement in their overall coherence in the inquiry test. Students from teachers A and B combined began with an overall coherence score of 56% (SD=36%) in the pre-test and improved to 66% in the post-test (SD=33%; \( t_{57} = 2.3, p = 0.023 \)). These gains are marginally lower than the 28% gains from the prior ThinkerTools trial (\( t_{85} = 1.85, p = 0.07 \)) lending some support to our hypothesis that reducing the amount of self-assessment in the ThinkerTools curriculum, compared to the prior version, will lead to students being less able to do coherent research in which their question, hypotheses, experiments, results, and conclusions all fit together.

Summary and Discussion of Inquiry Results

Overall, these results suggest that METT was successful at teaching students about scientific inquiry and the gains in measures of inquiry for Teacher A's students were similar to those for the prior ThinkerTools curriculum. It is likely that this version and the prior version of the curriculum were successful for similar reasons. Both curricula scaffolded student learning about the inquiry cycle, and students in both versions of the curricula gained familiarity with the inquiry process by using this framework in each curricular module and in their project work in the last module of the curriculum.

Differences between the inquiry outcomes between the two instructional trials can probably be attributed to differences in the curricular approaches. For example, while there were
no differences in the gains scores for the conclusion subsection between teacher A’s students and those from Teacher C's students in the prior ThinkerTools trial, teacher A’s students' post-test scores were significantly higher than her students' post-test scores for the prior ThinkerTools curriculum. This may indicate that the METT approach was more successful at helping students form better conclusions, because they had a better idea of the form that a scientific model should take and of the criteria a good model should meet. Lower gains in the coherence subscore of the inquiry test for METT students may be attributed to a smaller emphasis on student self-assessment compared to the prior version, and White and Frederiksen (1998) demonstrated a relationship between self-assessment and coherent inquiry.

Finally, teacher B’s students underperformed teacher A’s students on this Inquiry Assessment with respect to their results and their conclusions subscores, which are the two most difficult sections of this inquiry test. These findings may be attributed to teacher B having covered less of the METT curriculum, and/or to his being a less-experienced teacher, and/or to his lack of experience teaching with ThinkerTools inquiry curricula. It seems reasonable that teacher preparation and experience, as well as the length of time students spend learning about inquiry, would affect the outcome of an inquiry-oriented curriculum like METT. In fact, it would have been surprising if the results had been otherwise.

**Physics Knowledge**

In this section of the paper, we address the question of whether METT affected students’ conceptual physics knowledge. One would expect that improving students’ modeling knowledge would lead to gains in their physics knowledge. In particular, we hypothesized that enabling students to understand the nature of models (what a model is), the process of modeling
A Model-Centered Approach

(how one creates and revises a model), as well as what how to evaluate models (what is a good model?), and the purpose of models (what are the models for?) would enable students to create or recognize more accurate and useful models and be able to use them for generating predictions and explanations. For example, students with a better understanding of our four modeling dimensions might have a better understanding of the model form they were creating (a rule that predicts and explains), how to develop and refine it (the model needs to be accurate and plausible), and how to apply it – particularly on far transfer problems that involve applying physics concepts in entirely new contexts.

Did students show evidence of improvement of their physics knowledge? And, if so, how do these results compare with those from the prior version of the curriculum?

The Applied Physics Test

In determining students’ physics knowledge, we used the Applied Physics Test, a paper and pencil assessment that was administered before and after the curriculum. The Applied Physics Test is similar to the one used in the prior version of the ThinkerTools curriculum, and is comprised of twenty-two enhanced multiple choice questions designed to assess students’ understanding of conceptual models of force and motion. The assessment items were derived from prior research on students’ force-and-motion misconceptions and have been used extensively in earlier ThinkerTools research. For further description of the assessment items, see White (1984, 1993a), White and Frederiksen (1998), and Schwarz (1998).

Overall Results

Analyses of 128 students from six METT classes (four teacher A classes and two teacher B classes) indicates that students showed an improvement in their physics knowledge on items
whose concepts were addressed in METT, with a mean of 45% correct on the pre-test and 61% on the post-test ($t_{127} = 8.04, p < .001, η = 1.04$). The twelve items on this test included no transfer, near transfer, far transfer, and beyond scope$^2$ problems involving one and two-dimensional motion and collisions in one-dimension (See table 9 for results). In order to compare the performance of METT students with the performance of students who participated in the prior version of the curriculum, we calculated another score based on a subset of nine questions from the Applied Physics Test that were also included in the physics test used to evaluate the prior version of the curriculum (problems about two-dimensional motion and collisions in one-dimension). The students from the METT curriculum showed a 8.6% gain, with a pre-test mean of 43% correct and a post-test mean of 52% correct ($t_{124} = 3.66, p < .001, η = .43$). The students from the prior version of the curriculum showed a gain of 15.3%, with a pre-test mean of 44.6% and a post-test mean of 60% ($t_{97} = 6.26, p < .001, η = .88$). We should note that the smaller gains ($t_{221} = 1.88, p = .06$) for students in the METT curriculum are entirely due to their lower gains in the area of two-dimensional motion. Unfortunately, there were no items in common with the prior trial for the topic of one-dimensional motion, either with or without friction. Additionally, METT students did as well as or better than students in the prior study on the beyond scope collision problems – whose topic was not covered by either curriculum.

Insert table 9 about here

In carrying out our repeated measures analysis of variance on the Applied Physics Test
scores, we found that students’ academic achievement level was again a significant factor. Students with lower academic achievement on standardized test scores showed smaller gains on the physics test (a mean gain of 2%) than did students with higher academic achievement (with a mean gain of 15%), $F(1, 111) = 7.02, p < .01$.

**Analyses of Individual Items**

Analysis of individual items on the physics assessment gives further information about how students improved their understanding of the physics from METT. See table 9 for details summarizing item performance and Schwarz (1998) for a complete description of the data.

**One-Dimensional Motion**

METT students showed significant improvement on two of the three questions about one-dimensional motion, which is a topic that is addressed primarily in the first and second module of the curriculum. The third question on the topic exhibited ceiling effects. All of these problems were no or near transfer problems. For example, students were asked:

Imagine that a spaceship is coasting along in deep space. It is not near any planets or other outside forces. What will be true about the speed of the spaceship as it moves along?

A. the speed will decrease

B. the speed will remain the same

C. the speed will increase

Sixty-three percent of the students chose the Newtonian answer B on the pre-test compared to 82% on the post-test ($\chi^2 (1, n=125) = 13.7, p < .001$). The remaining incorrect responses in the post-test were evenly split between A and C. Along with the other two one-dimensional motion
questions, we cannot compare these results to those from the prior instructional trial, because these items were not included on the previous ThinkerTools physics assessment.

**Two-Dimensional Motion**

Students significantly improved their performance on four of the six two-dimensional motion questions, the topic primarily addressed in the third module of METT. For example, students significantly improved their responses to the question in figure 6, which is a near-transfer problem.

Eighteen percent of students chose the Newtonian response ‘B’ to the question in the pre-test compared to 42% in the post-test ($\chi^2 (1, n=110) = 16.95, p < .001$). In comparison, 22% of students in the prior ThinkerTools curriculum$^3$ chose the Newtonian response in the pre-test compared to 70% in the post test ($\chi^2 (1, n=79) = 30.08, p < .001$). Less exposure to the Newtonian computer microworld combined with exposure to the complex model-design rules provided by the Model-Enhanced ThinkerTools software may have contributed to the lower level of improvement on such items for METT students.

METT students did not improve their responses to the remaining two-dimensional problems including one near-transfer problem in which they were asked how to hit a hockey puck traveling diagonally in order to make it travel straight, and the other a far transfer problem in which they were asked to compare the time it would take to cross a river in a boat with and without a perpendicular current. Again, complex model-design rules for this topic and less exposure to the
Newtonian microworld may have made it difficult for students to develop Newtonian conceptual models that can be used to generate the "correct" answers to these problems.

One-Dimensional Collision Problems

Finally, METT students showed improvement on two of the four beyond scope problems relating to force and motion in the context of one-dimensional collisions. The other two problems in this area exhibited ceiling effects. On one of the problems for which there was a significant improvement, students were told that a moving billiard ball collides with a non-moving billiard ball on a pool table. They were then asked whether the non-moving billiard ball exerts a force on the original moving ball, and if so, in what direction that force is exerted. Prior to the METT curriculum, 34% of students chose the Newtonian response, that the force exerted was in the opposite direction to the motion of the moving billiard ball, compared to 50% in the post-test ($\chi^2 (1, n=116) = 8.805, p = .003$). By contrast, thirty percent of students in the prior ThinkerTools trial chose the Newtonian response in the pre-test compared to 34% in the post-test, showing no significant gains on this problem ($\chi^2 (1, n=96) = .5, p = .48$).

Summary and Discussion of Physics Results

Analysis of students’ performance on the physics assessment indicates that METT fostered improvement in students’ physics knowledge on almost all one- and two-dimensional problems. However, METT was not as successful as the prior ThinkerTools at promoting students’ understanding of two-dimensional motion. As previously mentioned, we suspect that this result may be due to several factors, including less exposure to the Newtonian computer models and reliance on real-world experiments for inducing the "correct" laws as well as to the complexity of the two-dimensional modeling options provided by the model-enhanced version of
the software. When students are learning about two-dimensional motion, interacting with Newtonian computer microworlds may be more effective at helping students visualize and internalize the normative physics than asking students to do real-world experiments and then reason through complicated computer-modeling rules.

Thus, we found that while the model-centered approach improved students’ physics performance, it produced lesser gains than the non-model enhanced version. This raises an interesting philosophical issue about the purpose of instruction. If the primary pedagogical goal is for students’ to develop a conceptual understanding of force and motion, then having students interact with carefully designed sequences of activities, set in the context of a Newtonian microworld, can achieve that goal without an added inquiry or modeling component, as prior research by the second author indicates (White, 1984, 1993). The issue then centers on deciding what science curricula and instruction should emphasize (White & Schwarz, 1999). Here we argue that developing expertise about modeling and inquiry, as well as the subject matter, should all be considered crucial components of good science education. The challenge then becomes developing an instructional approach in which the development of these areas mutually supports one another and enables students to successfully develop all of these components of scientific expertise. As indicated in the next section, METT is a step in that direction, but the fact that the prior ThinkerTools curriculum produced greater improvement on two-dimensional force-and-motion problems suggests that our model-centered approach needs some refinement.
RESULTS – CORRELATIONS AMONG MODELING, INQUIRY, AND PHYSICS EXPERTISE

In the prior sections of this paper, we presented results showing that METT improved students’ understanding of scientific modeling, inquiry skills, and physics knowledge, and we discussed how the two ThinkerTools curricular approaches elicited different outcomes. Here we present correlational evidence addressing our final research question, which asks how the development of modeling, inquiry, and physics expertise may be related to one another.

Our first hypothesis is that understanding the nature and purpose of models should foster the development of physics knowledge, given a curriculum like METT in which students learn about models as they engage in inquiry to develop and test models of physical laws. For instance, knowing the forms that models can take should help students in the development of their own conceptual models. Also, understanding that ThinkerTools models, and models in general, can be used for prediction and explanation should encourage students to use their models for prediction and explanation. At the same time, seeing that the computer model can follow alternative rules and that these rules lead to microworlds that behave differently may help students understand that in learning, they need to consider alternative explanatory models for the phenomenon being studied. In short, knowing about the nature and purpose of modeling may help students to come up with alternative conceptual models, evaluate those models, and utilize their best models to predict and explain phenomena.

Our second hypothesis is that modeling knowledge should play a reciprocal role in fostering the development of inquiry skills. On the one hand, knowing about the form and
evaluation of models should facilitate students’ investigatory and analytic skills. For instance, if students understand that their model needs to be in the form of a predictive rule, that understanding may help them formulate hypotheses, design experiments, and analyze their data. On the other hand, inquiry skills, such as knowing how to create hypotheses, conduct an investigation, analyze data, and draw a conclusion may help students understand the process of modeling. Thus we conjecture that modeling and inquiry should have a mutually supportive developmental relationship.

To investigate these hypotheses, we calculated correlations between students’ initial and acquired knowledge about modeling, inquiry, and physics. If our hypotheses are accurate, we would expect to see patterns of correlational evidence that support them. For instance, if modeling knowledge is important for learning inquiry and physics in the METT curriculum, then we would expect to see significant correlations between initial modeling knowledge and acquired inquiry knowledge, as well as between initial modeling knowledge and acquired physics knowledge. We would also expect to significant correlations between acquired modeling knowledge and acquired inquiry knowledge, as well as between acquired modeling knowledge and acquired physics knowledge. Further, we expect to see larger correlations between the post-test scores than the pre-test scores, because acquiring modeling knowledge contributes to fostering physics and inquiry knowledge and vice versa, when students learn using an integrated approach like that of METT, but not in their prior learning in which modeling, inquiry, and subject matter learning are not integrated.

The correlations among the three written assessments, taken before and after METT, are presented in Tables 10, 11, and 12. The assessments include: (1) the Modeling Assessment, (2)
the entire Applied Physics test, and (3) the Scientific Inquiry test. Performance on the post-tests includes knowledge of modeling, inquiry, and physics acquired from the METT curriculum. The pre-tests evaluate knowledge that has been acquired through the standard science curriculum, including the participating teachers’ prior curricula, as well as exposure to science outside of school. The sample is made up of the forty-six students whose inquiry pre- and post-tests were analyzed, which includes all students from teacher A and teacher B’s first two periods who had taken the pre-tests and post-tests for all of the assessments. Correlations from a larger sample of students, which includes data for students whose inquiry tests were not analyzed, show similar patterns for the most part and are reported elsewhere (Schwarz, 1998). We did not do separate analyses for each teacher because the sample sizes are too small for such a correlational analysis to have any power.

Note that in Table 10, the correlations contained in parentheses are corrected for unreliability of the column measures, and these should be used in comparing correlations within a row. In Tables 11 and 12, the correlations are corrected for unreliability of both the row and column variables. The reliabilities used in making these corrections (the coefficient alphas) are .76 for the Modeling pre-test, .81 for the Modeling post-test, .45 for the Applied Physics pre-test, .63 for the Applied Physics post-test, .90 for the Inquiry pre-test, and .88 for the Inquiry post-test. These reliability coefficients are an indication of the consistency of students' performance across the items within each test. The low reliability for the physics pre-test suggests that students' physics knowledge is item or situation specific, rather than based on a consistent set of principles used for solving the problems. The higher reliability of the physics post-test suggests that the METT curriculum is more successful than the students' prior science education
opportunities in creating a consistent conceptual base for solving the physics problems on the test.

Several interesting findings can be seen in this pattern of correlations. The first is that the modeling pre-test correlates significantly with both the physics and inquiry post-tests (see Table 10), supporting our hypothesis that modeling knowledge may play an important role in developing inquiry skills and physics knowledge. There is a similar finding for the Inquiry pre-test, suggesting that inquiry knowledge also plays a role in developing modeling and physics knowledge. Since inquiry expertise, as it is represented in the ThinkerTools curriculum, is centered on the development and testing of models, it is not unexpected to find that entering the curriculum with these kinds of knowledge and beliefs will enhance students’ learning in the METT curriculum.

In contrast, the Physics Pre-test does not correlate significantly with either the Modeling Post-test or Pre-test or with the Inquiry Post-test or Pre-test. This lack of correlation may be due to the nature of students' prior physics knowledge, which, based on the low reliability of the physics pre-test, appears to be item specific and lacks a coherent set of principles for solving the problems. Such knowledge appears to be less important as a base on which to build physics knowledge within the METT curriculum than does knowledge of modeling and of inquiry.

Another notable finding is that the modeling post-test correlates significantly with the other post-test measures (see Table 12). This result also supports our hypothesis that
developing modeling knowledge can play an important role in the acquisition of physics and inquiry knowledge. Further, a comparison of Tables 11 and 12 reveals that the modeling and physics post-test scores (see Table 12) are more highly correlated than are the modeling and physics pre-test scores (see Table 11). Similarly, the inquiry and physics post-test scores (see Table 12) are more highly correlated than are the inquiry and physics pre-test scores (see Table 11). Thus physics knowledge appears to show greater consistency and be more closely related to modeling and inquiry knowledge after the METT curriculum than before. In fact, as indicated in Table 12, all of the post-tests are significantly correlated with one-another, which suggests that, in the METT approach to science education, all three areas of knowledge undergo a mutual, interlinked development as the curriculum progresses.

These patterns of correlations lend support to our hypotheses that modeling knowledge can enhance the development of physics knowledge and inquiry skills and that modeling and inquiry expertise are related. While this Model-Enhanced version of the curriculum did not produce greater gains in inquiry and physics expertise than the prior version of the curriculum, the significant correlations of the Modeling Pre-test and Post-test with the Physics and Inquiry Post-tests suggest that modeling expertise and its acquisition are related to learning about physics and inquiry in the METT approach to science education. The data suggest that knowledge of the nature of models and modeling (what a model is and how models are constructed, revised, evaluated, and used) helps students to create conceptual models for understanding physics, to apply their models to different problem solving contexts, and to develop skills for scientific inquiry.
DISCUSSION AND CONCLUSIONS

Summary of Findings and Implications

Our findings indicate that a model-centered approach to science education can be effective at teaching students about scientific modeling, inquiry, and physics. For Teacher A's students, the METT curriculum resulted in a significant improvement in their performance on measures of modeling (a .5 $\sigma$ gain) and inquiry (a 1 $\sigma$ gain) expertise, which are notoriously difficult to teach, and a one sigma improvement in their performance on the physics test, for an area of physics known to be highly problematic for students (McDermott, 1984). Teacher B’s students improved significantly on the physics test, but not on the inquiry and modeling assessments. These differential results indicate that the impact of METT is teacher dependent, and may be explained by differences in teacher experience and curriculum use (Teacher B was a less experienced teacher and he also chose to omit portions of the curriculum).

Analysis of students' performance on both the Modeling Assessment and the Modeling Interview indicates that our pedagogical approach helped students learn about the nature and purpose of models. For instance, after completing the METT curriculum, many students were able to identify various abstract representations as models and understood how models can be used to predict and explain. Further, most students learned that scientific and computer models can be useful in a wide variety of ways, which include representing ideas, visualizing and predicting phenomena, testing hypotheses, and conducting investigations that are not otherwise possible. These results suggest that the students who participated in METT understood more about the nature and utility of modeling than the seventh graders surveyed by Grosslight et al. (1991).
Our findings also indicate that our pedagogical approach was less successful in promoting students’ understanding of model evaluation and revision. Many had limited understanding of how models are created or that scientists revise their models in the light of new insights or new data. In addition, while the students used model evaluation criteria during the curriculum, many gave post-test answers indicating that they considered all models to be of equal value. We speculate that the particular culture of the school, which emphasizes that everyone’s ideas must be respected and valued, may have contributed to the lack of improvement on test questions about the comparative value of alternative models.

Our finding that students learned more about the nature and utility of models and less about the evaluation and revision of models has several implications. First, it suggests that students’ modeling knowledge may not be adequately described in terms of a unidimensional developmental progression, such as that proposed by Grosslight et al. (1991). The idea that there are different components to understanding modeling, such as nature of models, model evaluation, revision, and model utility, may provide a more accurate and useful view of modeling expertise. And, as we have argued throughout this article, all of these are necessary components of a coherent, in-depth understanding of science.

Such a differential outcome may also indicate that further refinements to the METT curriculum and the Modeling Assessment are necessary. It is possible the curriculum needs to be refined to better foster students’ skills in evaluating and revising models. It is also possible that our Modeling Assessment did not provide a sensitive or accurate measure of these capabilities. The physics test gains indicate that students were capable of evaluating and revising their models. This suggests that the Modeling Assessment questions relating to model evaluation and revision
may not elicit the same responses (or be as accurate) as seeing how students evaluate and revise their models during their actual inquiries. We need to address all of these possibilities in our future research.

In addition to our findings that students improved their understanding and skills relating to modeling, inquiry and physics, a comparison of students' learning in these areas for METT versus the prior ThinkerTools curriculum suggests some interesting tradeoffs regarding these different curricular approaches. It appears that METT enables students to learn more about modeling, as much about inquiry, and somewhat less about the physics of force and motion than does the prior ThinkerTools curriculum. Although the Modeling Assessment was not available during the prior ThinkerTools study, subsequent interviews with ThinkerTools' students revealed that they lacked an understanding of the nature and purpose of models (Schwarz, 1996), which became our primary motivation for creating the METT curriculum. The gain on our measure of inquiry for METT was similar to that for the prior ThinkerTools curriculum. The one improvement in physics scores for the METT curriculum, though large, was not as great as that for the prior version of ThinkerTools, in which students interacted with only Newtonian computer simulations. These differential findings suggest that there is a partial tradeoff between enabling students to learn more about modeling, by creating their own non-Newtonian models, and enabling students to developing a normative understanding of physics through interacting only with Newtonian computer models.

An additional set of findings relates to the differential pre and post-test performance of Teacher A versus Teacher B students. For instance, Teacher B's students had higher scores on pre-test questions about the nature and purpose of modeling than did those of Teacher A, which
may be a result of his focus on modeling in an astronomy curriculum that he developed and
taught prior to teaching METT. Thus, Teacher B’s astronomy curriculum and his prior emphasis
on modeling appear to have been effective in developing students’ knowledge of modeling prior
to the METT curriculum.

Further, teacher B’s students did not show significant improvements on the Inquiry Test,
whereas students of Teacher A did. This finding has several possible explanations. Teacher B
included just two full modules of the METT curriculum and part of a third, while Teacher A
included the full four modules. Thus, in Teacher B’s implementation, students encountered only
two full experiences with the Inquiry Cycle, while Teacher A’s students had 4 full inquiry
experiences. Also, since Teacher B has a stronger background in the science content, he was
better equipped to provide students with guidance in developing the science content, but this also
may have undermined his students' learning through inquiry. Prior research suggests that for the
learning of physics to be dependent on the learning of inquiry, teachers must emphasize inquiry
and scaffold the inquiry processes of their students (White & Frederiksen, 1998). In other words,
they must rely on inquiry for developing their students' knowledge of the science content and
must refrain from engaging in direct instruction of the force-and-motion models that the students
are trying to develop. The first author observed that Teacher A placed more emphasis on inquiry
and less on science content than Teacher B. Having no formal training in science, Teacher A relied
on scaffolding the inquiry processes of her students to develop their knowledge of the science
content.

Finally, our correlational evidence also suggests that modeling knowledge can play an
important role in science learning. We found that prior knowledge of inquiry and modeling are
related to how much students learn from the METT model-centered approach to science education, but that prior knowledge of physics isn't related to how much they learn about physics, modeling, or inquiry. The patterns of correlations among pre-tests and post-tests also support our claim that modeling knowledge is useful for learning inquiry and physics. There was a significant correlation between initial modeling knowledge and acquired inquiry knowledge, and also between acquired modeling knowledge and acquired inquiry knowledge. Similarly, there was a significant correlation between initial modeling knowledge and acquired physics knowledge, and between acquired modeling knowledge and acquired physics knowledge. Thus, the correlational evidence suggests that knowledge of physics, modeling, and inquiry become interrelated in the METT approach and mutually support the development of one another.

Taken together these findings suggest that the METT approach, in which students engage in inquiry to develop and test models of domain phenomena, has the potential to enable learners to acquire a sophisticated and coherent understanding of the scientific enterprise. While we feel that there are aspects of the approach that need revision, the gain scores on our modeling, inquiry, and physics assessments as well as the significant correlations among the post-test scores are highly encouraging.

Further, our curricular approach shows how science curricula can combine model creation, evaluation, revision and use with opportunities for reflecting on the nature of models and the process of modeling. A modeling approach with these features can successfully promote an understanding of models, while curricula that have primarily focused on conceptual change without sustained reflection on the process of modeling have not (e.g. Wiser et al., 1988).
Given the potential benefits of learning about the nature and process of modeling, we propose that model-centered instruction should be incorporated more widely into science curricula. Students would benefit from being introduced to multiple types of models and from having many opportunities to create, revise, and reflect on models. This exposure should provide students with an opportunity to understand a wide variety of models as well as the subtle complexities involved in the process of modeling (Collins & Ferguson, 1993). Such a sustained emphasis on models during science class can potentially have a much greater impact on students understanding of the nature and purpose of models and of model creation and evaluation than is feasible in a ten and a half week curriculum.

Challenges for Future Research and Implementation

While this study has shown that a model-centered curriculum and instruction can benefit students’ learning, we experienced several difficulties in fostering students’ modeling expertise. One problem is that some students in METT did not appear to understand why they were engaged in the process of scientific modeling. Other researchers promoting model-centered instruction have also experienced this difficulty (Brand et al., 1998; Lewis, Rader, Brand, & Carlone 1997; Barowy & Roberts, 1999). Certainly, including additional class discussions about why students are engaged in modeling might address this difficulty. However, enabling students to see the benefits of modeling in pursuing their own interests and ideas would probably have a greater impact. Modeling curricula should better fit the interests of young students, as modeling needs to have both purpose and payoff. For instance, the ThinkerTools research group has developed a curriculum module in which students investigate the impact of listening to music on how well they do warm-up exercises (typically a short problem solving or writing task) at the
beginning of class. The payoff for students is that, if their research provides support for the benefits (or at least lack of harm) due to music, then the teacher allows students to listen to music during warm-up exercises.

A further challenge is that students had difficulty in understanding the process of modeling, despite our model-centered approach. This finding may suggest that choosing from alternative possible models, as students do with our model-enhanced ThinkerTools software, is not as effective as creating models from scratch. Thus students may better understand the processes of model creation and evaluation if they work with more open-ended modeling environments, such as the model-expressive programs listed in figure 1, which enable them to create and modify models.

We also face challenges in redesigning our approach to better meet the needs of both high and low academic achieving students, as the prior ThinkerTools met with greater success in reducing the disparity between high and low achieving students. We suspect several factors may have contributed to the METT results. First, METT included less self-assessment than the prior ThinkerTools, and engaging in self-assessment was found to be a significant contribution to the learning of lower-achieving students (White & Frederiksen, 1998). Secondly, METT students spent less time interacting with the Newtonian microworlds, because they were free to choose non-Newtonian models, which may have reduced their ability to develop conceptual models that embody Newton's laws of motion. Finally, the epistemological and metaconceptual ideas about modeling that are addressed in METT are challenging. Possibly providing more self-assessment activities, increased levels of scaffolding in the model-design process, and more autonomy (so
that students can choose topics that interest them) might address some of the disparity between high and low achieving students.

In addition to addressing these difficulties, several other practical issues remain if we are to successfully promote and implement a model-centered approach to science education. Teachers need to have an understanding of the nature of models and the process of modeling so that they can help their students learn about modeling. Also, they must be able to give students enough time in their curricula to engage in scientific inquiry and modeling. State content science standards, such as those in California, require large amounts of content coverage, making time for inquiry and modeling increasingly difficult. Finally, knowledge about modeling is not currently valued or assessed on standardized tests.

Extensions of This Research

In this paper, we provided theoretical and empirical evidence that modeling knowledge is an important component of learning science. While our efforts to develop this area of students' knowledge were reasonably successful, we continue to work at improving this approach to make it successful for all students. We are pursuing three different directions in addressing this goal.

We revised METT to improve students’ physics gains, while continuing to teach modeling knowledge. In one revised version, students begin by interacting with the Newtonian model and then gradually start using the Model-Design aspect of the software in later curriculum modules. In another proposed revision, students would interact with the Model-Design software for the simple conceptual models, like friction, and would interact only with Newtonian models for more counterintuitive and complex models, like friction-free or two-dimensional motion. In this way, students would be more likely to acquire a normative understanding of force and
motion, while still having an opportunity to choose from alternative computer models for phenomena where they are more likely to discover the normative models.

In addressing the need to develop teachers’ knowledge of the nature of models, the first author has used a model-centered approach in teaching an undergraduate general science class to pre-service elementary teachers. In the course, reflection about the nature of models occurred in class discussion and in reflective journals. The study found that these future teachers successfully learned about the nature of models and the process of modeling while creating, using, and revising physical models in astronomy and computer models in physics. However, the study also found that the different learning contexts and uses of models in the class elicited a different kind of model understanding. Unlike the METT students, who thought of models as abstract, more or less equivalent in value to one another, and useful for testing and exploring their own theories, these pre-service teachers thought of models as concrete or visual representations, more or less valuable depending on their accuracy and limitations, and useful for visualizing phenomena and for learning and teaching the correct scientific knowledge (Schwarz, 2002). The first author continues to investigate ways to enable pre-service teachers to learn about models and modeling tools so that they can incorporate them into their science teaching.

Finally, to enable students to learn more about the processes of scientific modeling and inquiry, the second author is creating an inquiry support environment, which includes a community of personified software advisors (Eslinger, White, & Frederiksen, 2001; Shimoda, White, & Frederiksen, 2002; White, Shimoda, and Frederiksen, 1999). In this environment, there are advisors for each step in the inquiry cycle, including a Hypothesizer, Investigator and Modeler, as well as general purpose advisors, including an Inventor, Planner, Monitor, and
A Model-Centered Approach

Reviser. The Modeler, for instance, provides students with information about different types of models and their uses, strategies and plans for developing models, and criteria for evaluating them. The Modeler also provides specific examples of all of these aspects of modeling in a variety of contexts (physics, biology, and social science). Students are guided by these advisors when they undertake research projects, such as when they do their ThinkerTools force-and-motion or their science fair projects.

Conclusions

We have argued that students should know about the nature of models and the process of modeling in order to be scientifically literate. Our research suggests that a model-centered approach can promote an understanding of models, and that understanding models and inquiry skills can support one another and the learning of physics. Our conclusion is that science instruction should couple inquiry learning with a coherent model-centered approach, which includes model creation, use, evaluation, and revision, as well as reflection about the nature of models and the role that different types of models play in scientific theorizing. Further advances in scientific modeling and computer technology will enable us to use a model-centered approach in various interesting and powerful ways, such as using multi-agent systems to permit students to create and test models of their own collaborative learning processes (White, Frederiksen, Frederiksen, Collins, Eslinger, and Loper, 2002).

Nonetheless, difficulties with model-centered approaches remain. As this research indicates, enabling students to acquire knowledge about modeling is difficult and requires time, effort, and teacher expertise. We have argued that it is important for students to learn about the processes and products of science. However, achieving this goal requires that educators rethink
their priorities for science education. Should content coverage be reduced to make room for learning about scientific inquiry processes and about scientific models? Our finding suggest that learning about modeling and inquiry can facilitate learning the science content, however, one has to reduce expectations for content coverage to achieve this goal.

Future research in this area will serve to advance our understanding and implementation of model-centered approaches. In particular, we continue to theorize about how modeling knowledge develops and how it may affect learning, and we are investigating the context-specificity of modeling knowledge, how it may be applied in novel contexts, and whether there is a developmental progression for modeling knowledge (Schwarz, 2002b). Additionally, we believe that it will be important to determine which aspects of this model-centered approach are necessary and sufficient to provide an epistemologically-rich environment for student learning. Finally, it will be critical to develop good approaches to fostering teachers’ understandings and use of model-centered approaches in science education, so that students may learn using such approaches.

Teaching students about models and the process of modeling is a challenging yet promising way to develop their understanding of many crucial aspects of science. As a result, we need to continue refining model-centered approaches by investigating authentic and motivating ways of incorporating modeling into curricula, by developing innovative pedagogical approaches that combine model expression and exploration, and by creating and experimenting with different kinds of computer-based modeling environments that enable students to work with a wide variety of models. In addition, we need to engage teachers in learning about the modeling process as we refine our theories about how modeling knowledge develops and discover which aspects of
modeling are critical to successful learning environments. Work in these areas should lead to greater benefits for students and teachers alike, as well as informing our understanding of scientific modeling and the role that modeling can play in the learning of science.

ACKNOWLEDGEMENTS

This research was supported by a National Science Foundation science and design fellowship for the first author and by grants to the second author from the National Science Foundation (Awards DPE-8400280 and REC-0087583) and the James S. McDonnell Foundation (Award 9135). We gratefully acknowledge the contributions of members of the ThinkerTools research team and other graduate students at U.C. Berkeley for their valuable input regarding the research. In particular, we thank John Frederiksen for his advice on research design and analysis and Christopher Schneider for his expertise in designing and implementing the software. We are also extremely grateful to the middle school teachers and students who participated in this project. Finally, we thank the two anonymous reviewers for their comments on an earlier draft of this article.

REFERENCES


Shimoda, T., White, B., & Frederiksen, J. (2002). Student reflective inquiry: Increasing levels of agency through modifiable software advisors. Science Education, 86, 244-263.


Table 1
Types of Modeling Knowledge

1. **Nature of models:**
   - **Kinds of models and model attributes:** What is a model?
   - **Model content:** What do models represent?
   - **Multiple models:** Can there be different models for the same object/phenomena?
   - **Constructed nature of models:** Do models represent absolute reality?

2. **Nature of modeling:**
   - **Modeling process:** What is involved in the modeling process?
   - **Designing and creating models:** How are models constructed?
   - **Changing models:** Would a scientist ever change a model?

3. **Evaluation of models:**
   - **Model evaluation:** Is there a way to decide whether one model is better than another?
   - **Model criteria:** What kinds of criteria are used to evaluate models?

4. **Purpose/Utility of models:**
   - **Purposes of models:** What are models for?
   - **Utility of models in science and science classes:** How can models be useful for scientists or students in science classes?
   - **Utility of multiple models:** What is the purpose of having multiple models of the same phenomena/object?
## Table 2
Overall Results of Student Model Understanding

<table>
<thead>
<tr>
<th>Model Understanding</th>
<th>Results from the pre/post Modeling Assessment</th>
<th>Results from the Modeling Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Results (including all dimensions of modeling knowledge)</td>
<td>64% pre-test, 71% post-test* for six seventh-grade classes on eighteen items. Significant gains for teacher A students but not for teacher B students. Academic achievement also a significant factor.</td>
<td>Evidence of sophistication in the various dimensions. Evidence that knowledge was retained after the curriculum and instruction.</td>
</tr>
<tr>
<td>I. Nature of models dimension</td>
<td>67% pre-test, 73% post-test* on seven items. Significant gain for teacher A students but not teacher B students, partly due to ceiling effects on several items.</td>
<td>Evidence of sophisticated ideas about the nature of models, multiple models, and the constructed nature of models. Teacher A students did better on some questions and teacher B students did better on others.</td>
</tr>
<tr>
<td>II. Nature of modeling dimension</td>
<td>52% pre-test, 56% post-test on three items.</td>
<td>Moderate sophistication about model revision.</td>
</tr>
<tr>
<td>III. Evaluation of models dimension</td>
<td>52% pre-test and post-test on three items.</td>
<td>Mixed results. Students demonstrated strong use of model evaluation criteria, but were relativistic about &quot;is any model as good as another?&quot;</td>
</tr>
<tr>
<td>IV. Purpose of models dimension</td>
<td>63% pre-test and 68% post-test on five items. Significant gain for teacher A students but not teacher B students, partly due to ceiling effects on several items.</td>
<td>Evidence of highly sophisticated ideas about the purpose of the model-design software and the general purpose of modeling.</td>
</tr>
</tbody>
</table>

* significant to p < .001 level
Table 3
Frequencies of Interview Responses for the Definition of a Model

<table>
<thead>
<tr>
<th>Codes for Definition of a Model</th>
<th>Comment Codes*</th>
<th>Level Rating**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview Questions: “In this context, what do we mean by a model? … Now that we’ve talked a little bit, can you tell me what you think scientific models are in general?”</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A’s (strong)</strong> A strong response indicates an abstract notion of a model as something that helps predict or explain and is not necessarily visual or concrete.</td>
<td>16 (73%)</td>
<td></td>
</tr>
<tr>
<td>It’s:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. a rule (that predicts and explains)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2. a representation (that predicts and explains)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3. a theory or idea (that predicts and explains)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4. a reasoning structure (like a computer model which will fit your hypothesis or experiment) (that predicts and explains)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5. a summarizing conclusion if it predicts and explains</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Properties (These do not count in the overall categorization):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. it doesn’t necessarily have to be a physical object; it can be a mental object or words</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7. it should be accurate, plausible, useful, and consistent</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8. it is often simplified</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>B’s (moderate)</strong> A moderate response is a two or three-dimensional visual model.</td>
<td>3 (14%)</td>
<td></td>
</tr>
<tr>
<td>It’s:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. a diagram or design</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2. a picture or drawing</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3. a 3-dimensional object</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Properties (These do not count in the level ratings):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. it’s useful</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5. you can see it visually</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6. it shows you something</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>7. to show how things are made</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8. it helps construct objects</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>C’s (weak)</strong> A weak response is one in which the student thinks of a model as a concrete object or something that is not a model.</td>
<td>1 (5%)</td>
<td></td>
</tr>
<tr>
<td>1. a conclusion if it is a statement and not an abstraction that predicts or explains</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. a smaller copy/version of an object</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3. to do it in the real world</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4. something scientific</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5. showing your work or what you learned</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6. a scientific gadget</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>NA</strong> Other/not clear: No information, not enough information, or interviewer did not ask the question</td>
<td>2 (9%)</td>
<td></td>
</tr>
</tbody>
</table>

*Note. The frequencies include all of the students’ responses and may total more than that number of students.
**The overall level rating is the highest rated response the student gave to the question; they can be added to equal the total number of students.
Table 4
Frequencies of Interview Responses for Model Revision

<table>
<thead>
<tr>
<th>Codes for Model Revision</th>
<th>Comment Codes*</th>
<th>Level Rating**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A’s (strong)</strong> A strong response may capture the sense that model revision may occur by rethinking data and its implications as well as the purpose of the model</td>
<td>6 + (1) = 7</td>
<td>(32%)</td>
</tr>
<tr>
<td>1. They are changed in order to improve our interpretation of reality (expert response from Grosslight et al.)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2. A scientist could change their model if they looked over their data again and re-thought its implications</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. A scientist could change their model if they thought more about it. (This could also happen by discussing it with other scientists)</td>
<td>6 + (1) = 7</td>
<td></td>
</tr>
<tr>
<td><strong>B’s (moderate)</strong> A moderate response captures the sense model revision occurs because scientists gain new information or because they have made mistakes.</td>
<td>3 + (8) = 11</td>
<td>(50%)</td>
</tr>
<tr>
<td>1. Scientists could change their model to make it more accurate if they found out new information or evidence (by conducting more investigations)</td>
<td>6 + (7) = 13</td>
<td></td>
</tr>
<tr>
<td>2. Scientists could change their model if they made mistakes during the experimentation</td>
<td>3 + (3) = 6</td>
<td></td>
</tr>
<tr>
<td><strong>C’s (weak)</strong> A weak response captures the sense that model revision might occur for only vague reasons of correctness or might not occur at all.</td>
<td>1 + (3) = 4</td>
<td>(18%)</td>
</tr>
<tr>
<td>1. A scientist would change their model if it wasn’t right; if something was ‘wrong’; to make sure it’s right</td>
<td>2 + (5) = 7</td>
<td></td>
</tr>
<tr>
<td>2. Scientists don’t change their models; They might put them aside for later, but they never throw out the chance to find out something about them.</td>
<td>(1)</td>
<td></td>
</tr>
</tbody>
</table>

Codes for students who were not asked the last interview question “Might scientists change their models even without more experimental data?” have been included in parentheses.

*Note. The frequencies include the addition of all students’ responses and may total more than that number of students.

**The overall level rating is the highest rated response the student gave to the questions; they can be added to equal the total number of students.
Table 5
Frequencies of Interview Responses for Model Evaluation Criteria

<table>
<thead>
<tr>
<th>Codes for Model Evaluation Criteria</th>
<th>Comment Codes*</th>
<th>Level Rating**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codes related to how students evaluated their research findings and what criteria they used in the context of discussing their research project. Codes also based on how students evaluated their model in their thought experiment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A’s (good)</strong> A good response indicates that the student evaluated her model with scientific criteria or compared her model to other models.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Determining whether the model was accurate, plausible, consistent, and/or useful</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>2. Comparing the model to the Newtonian one or to other’s models and/or to the hypotheses</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3. Checking to see if the data or model make sense; thinking about how their data worked</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 (50%)</td>
<td></td>
</tr>
<tr>
<td><strong>B’s (medium)</strong> A medium response indicates the student evaluated his model for vague notions of correctness or clarity or used the evaluation criteria/sheet.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Evaluating oneself with the evaluation criteria/sheet</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. Determining whether the model was clear</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3. Remembering that the evaluation criteria were accuracy, plausibility, utility, or consistency</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (32%)</td>
<td></td>
</tr>
<tr>
<td><strong>C’s (weak)</strong> A weak response indicates the student evaluated her work by checking it or some other vague notion.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Checking your work (vague); Seeing whether the model is okay/good.</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>2. Showing the work to the teacher</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3. Making sure everything is complete</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4. Summing it up</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5. Writing everything down, and talking about results</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6. Evaluating yourself, others, your project (vague)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 (18%)</td>
<td></td>
</tr>
<tr>
<td><strong>NA</strong> Other/not clear: No information, not enough information, or interviewer did not ask the question</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note. The frequencies include the addition of all students’ responses and may total more than that number of students.

**The level rating for each question is the highest rated response the student gave to the question; they can be added to equal the total number of students.
### Codes for General Evaluation of Models

**Interview Question:** “In general, is any model as good as another?”

<table>
<thead>
<tr>
<th>Comment</th>
<th>Level Rating**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A’s (strong)</strong> A strong response is one that acknowledges that some models are better able to predict and explain phenomena than others.</td>
<td>8 (36%)</td>
</tr>
<tr>
<td>1. No. Some are less valid, accurate, plausible, well-thought out, or supported by evidence than others</td>
<td>6</td>
</tr>
<tr>
<td>2. No. Some are totally weird or complicated</td>
<td>3</td>
</tr>
<tr>
<td><strong>B’s (moderate)</strong> A moderate response indicates that all models are of equal value only when evidence can’t distinguish among them.</td>
<td>2 (9%)</td>
</tr>
<tr>
<td>1. Yes. When there is no way to know which one is right</td>
<td>2</td>
</tr>
<tr>
<td>2. Yes, until somebody’s tried it</td>
<td>1</td>
</tr>
<tr>
<td><strong>C’s (weak)</strong> A weak response indicates that there is no real way to judge a model from any other. They all have value.</td>
<td>10 (45%)</td>
</tr>
<tr>
<td>1. Yes, because everyone thinks differently</td>
<td>1</td>
</tr>
<tr>
<td>2. Yes, everybody has their own opinion</td>
<td>1</td>
</tr>
<tr>
<td>3. Yes, because it’s showing a point of view</td>
<td>1</td>
</tr>
<tr>
<td>4. Yes, because there is no right or wrong answers. Some have more of a probability of being right because of an experiment, but [nothing is ever really wrong.]</td>
<td>1</td>
</tr>
<tr>
<td>5. Yes, although it depends on how careful a person was.</td>
<td>1</td>
</tr>
<tr>
<td>6. Yes, as long as the person worked hard on it OR because everyone put hard work into making their model</td>
<td>3</td>
</tr>
<tr>
<td>7. Yes, because models aren’t answers; they’re possibilities</td>
<td>1</td>
</tr>
<tr>
<td>8. Yes. It may be wrong, but it’s your idea</td>
<td>1</td>
</tr>
<tr>
<td><strong>NA</strong> Other/not clear; No information, not enough information, or interviewer did not ask the question</td>
<td>2</td>
</tr>
</tbody>
</table>

*Note. The frequencies include the addition of all students’ responses and may total more than that number of students.

**The level rating is the highest rated student response to the question; they can be added to equal the total number of students.
Table 7
Frequencies of Interview Responses for the
General Purpose of Modeling

<table>
<thead>
<tr>
<th>Codes for General Purpose of Modeling</th>
<th>Comment Codes*</th>
<th>Level Rating**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview Questions: “What do you think computer or scientific models can be useful for? How do you think computer models can be useful for scientists? How do you think computer models can be useful for learning about science?”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A’s (strong) *A strong response indicates that models may be useful for envisioning, predicting, testing, or teaching.*

1. to visualize one’s own and other people’s models (represent ideas) 9
2. to test theories/make sure what you think is right (come to logical conclusions); compare models 8
3. to manipulate models 1
4. for predicting phenomena 8
5. to test or investigate things not otherwise possible in the real world (including things too complicated to see in the real world) 7
6. to help explain and show people what, why, or how something happens (about one’s theories) 6
7. to help figure out/understand what, why, or how something happens 10
8. to help people create better models 1
9. to teach people about the subject material of the model 6
10. to teach students how to experiment, collect data and write out models 1

18 (82%)

B’s (moderate) *A moderate response indicates a vague response that models may be useful for constructing an object, to get information, or to make life easier.*

1. for constructing/making an object 7
2. to get information 9
3. to make things more accurate and easier than in real life; for convenience 4

3 (14%)

C’s (weak) *A weak response indicates the student doesn’t think that models can be useful.*

1. they don’t really help much because the computer just follows the ideas you put into it

0 (0%)

NA *Other/not clear: No information, not enough information, or interviewer did not ask the question*

1 (5%)

*Note. The frequencies include the addition of all students’ responses and may total more than that number of students.

**The level rating of the student is the highest rated response the student gave to the questions, and so the overall level ratings can be added to equal the total number of students.
Table 8
Scientific Inquiry Assessment Scores

<table>
<thead>
<tr>
<th></th>
<th>METT Teacher A Pre/Post Tests</th>
<th>METT Teacher B Pre/Post Tests</th>
<th>Prior ThinkerTools Teacher C Pre/Post Tests</th>
<th>Prior ThinkerTools Teacher A Post-Test only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Inquiry Scores</td>
<td>47% (21%)</td>
<td>62%** (18%)</td>
<td>43% (29%)</td>
<td>49% (27%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40% (24%)</td>
<td>61%** (17%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40% (24%)</td>
<td>NA (21%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>56% (21%)</td>
<td></td>
</tr>
<tr>
<td>Hypothesis Sub Score</td>
<td>71% (19%)</td>
<td>73% (24%)</td>
<td>74% (27%)</td>
<td>73% (26%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>51% (25%)</td>
<td>64%** (22%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA (21%)</td>
<td>67% (21%)</td>
</tr>
<tr>
<td>Experiment Sub Score</td>
<td>43% (20%)</td>
<td>54%** (15%)</td>
<td>35% (26%)</td>
<td>41% (23%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40% (25%)</td>
<td>56%** (16%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA (18%)</td>
<td>51% (18%)</td>
</tr>
<tr>
<td>Results Sub Score</td>
<td>30% (23%)</td>
<td>62%** (19%)</td>
<td>26% (24%)</td>
<td>43%** (24%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30% (25%)</td>
<td>59% ** (24%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA (29%)</td>
<td>58% (29%)</td>
</tr>
<tr>
<td>Conclusion Sub Score</td>
<td>38% (32%)</td>
<td>59%** (27%)</td>
<td>41% (37%)</td>
<td>44% (31%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>34% (29%)</td>
<td>55%** (24%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA (33%)</td>
<td>43% (33%)</td>
</tr>
<tr>
<td>Coherence Sub Score</td>
<td>57% (36%)</td>
<td>70%* (29%)</td>
<td>55% (38%)</td>
<td>59% (39%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50% (33%)</td>
<td>77%** (27%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA (33%)</td>
<td>66% (33%)</td>
</tr>
</tbody>
</table>

Numbers in parentheses include the standard deviations for the means
** p < .01. *p < .05.
### Table 9
**General Physics Assessment Scores**

<table>
<thead>
<tr>
<th></th>
<th>METT Pre-Test</th>
<th>METT Post-Test</th>
<th>Prior ThinkerTools Pre-Test</th>
<th>Prior ThinkerTools Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Physics Scores on all items covered by METT</td>
<td>45% (15%)</td>
<td>61%** (20%)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Total Physics Scores on subset of items covered by</td>
<td>43% (19.9%)</td>
<td>52%** (23.1%)</td>
<td>44.6% (17.5%)</td>
<td>60%** (18.8%)</td>
</tr>
<tr>
<td>both METT and Prior TT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D without friction item (NT)</td>
<td>63%</td>
<td>82%**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1D with friction item (NoT)</td>
<td>91%</td>
<td>91%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1D with friction item (NoT)</td>
<td>75%</td>
<td>90%**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2D Motion item (NT)</td>
<td>18%</td>
<td>42%**</td>
<td>22%</td>
<td>70%**</td>
</tr>
<tr>
<td>2D Motion item (NT)</td>
<td>14%</td>
<td>16%</td>
<td>13%</td>
<td>49%**</td>
</tr>
<tr>
<td>2D Motion item (NT)</td>
<td>23%</td>
<td>47%**</td>
<td>33%</td>
<td>64%**</td>
</tr>
<tr>
<td>2D Motion item (NT)</td>
<td>12%</td>
<td>28%**</td>
<td>21%</td>
<td>58%**</td>
</tr>
<tr>
<td>2D Motion item (FT)</td>
<td>26%</td>
<td>39%**</td>
<td>30%</td>
<td>44%**</td>
</tr>
<tr>
<td>2D Motion item (FT)</td>
<td>11%</td>
<td>7%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1D collisions item (BSc)</td>
<td>94%</td>
<td>97%*</td>
<td>97%</td>
<td>97%</td>
</tr>
<tr>
<td>1D collisions item (BSc)</td>
<td>95%</td>
<td>95%</td>
<td>92%</td>
<td>93%</td>
</tr>
<tr>
<td>1D collisions item (BSc)</td>
<td>45%</td>
<td>52%</td>
<td>34%</td>
<td>41%</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1D collisions item (BSc)</td>
<td>34%</td>
<td>50%**</td>
<td>30%</td>
<td>34%</td>
</tr>
</tbody>
</table>

Numbers in parentheses include the standard deviations for the means. NoT stands for no transfer problems, NT for near transfer problems, FT for far transfer problems, and BSc for beyond scope problems.

** $p < .05$. *$p < .10$. 
<table>
<thead>
<tr>
<th>Learning Measures</th>
<th>Modeling Pre-test</th>
<th>Physics Pre-test</th>
<th>Inquiry Pre-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling Post-test</td>
<td>.65** (.83)</td>
<td>.08 (.12)</td>
<td>.44** (.52)</td>
</tr>
<tr>
<td>Physics Post-test</td>
<td>.40** (.58)</td>
<td>.28 (.53)</td>
<td>.45** (.60)</td>
</tr>
<tr>
<td>Inquiry Post-test</td>
<td>.36* (.49)</td>
<td>.14 (.22)</td>
<td>.69** (.77)</td>
</tr>
</tbody>
</table>

** p < .01. *p < .05. n = 46. Note: the correlations in parentheses are corrected for unreliability of the column measures.
<table>
<thead>
<tr>
<th>Learning Measures</th>
<th>Modeling Pre-test</th>
<th>Physics Pre-test</th>
<th>Inquiry Pre-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling Pre-test</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Physics Pre-test</td>
<td>-.05 (-.09)</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>Inquiry Pre-test</td>
<td>.42** (.51)</td>
<td>.17 (.27)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

** p < .01. *p < .05. n = 46. Note: the correlations in parentheses are corrected for unreliability of the column measures.
Table 12
Correlations of Post-tests

<table>
<thead>
<tr>
<th>Learning Measures</th>
<th>Modeling Post-test</th>
<th>Physics Post-test</th>
<th>Inquiry Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling Post-test</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Physics Post-test</td>
<td>.43** (.61)</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>Inquiry Post-test</td>
<td>.41** (.49)</td>
<td>.44** (.59)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

** p < .01. *p < .05. n = 46. Note: the correlations in parentheses are corrected for unreliability of the column measures.
Figure 1: This figure illustrates a range of modeling software placed on a continuum ranging from those that are more model-expressive to those that are more model-exploratory.
Figure 2: These screen shots illustrate how the software enables students to choose from among alternative laws of motion.
Figure 3: Three screen shots from a simulation activity in which students run an experiment and see the implications of the laws of motion they have selected.
Figure 4. The ThinkerTools scientific inquiry cycle.⁵
Accuracy: One way for scientists to judge whether a model is good is to see if it accurately describes the experimental data. In other words, does the model predict what actually happened?

Evaluation Question 1:

Look back at your model *The effect of friction on the speed of an object* (on page 6) and rate it for accuracy. Does this model describe and predict what actually happened? Think back to your data in order to help you answer this question.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>inaccurate</td>
<td>somewhat accurate</td>
<td>very accurate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why did you give your model this rating?

----------------------------------------------

Figure 5: Sample model evaluation instructions
Refer back to your *Motion with no forces* model and use it to answer the following question.

**Application Question 2.** You are playing hockey on a top-secret experimental ice rink. The rink is one mile long and the surface is so perfect that it actually has no friction. Your teammate (who has the puck) is on the other end of the rink from you. If she can get the puck to you, you can hit it into your opponent’s net (which is right next to you) and score the winning point! She decides to hit the puck to you. What does your model say should happen to the puck? (circle one)

(a) No problem! It will make it the whole length of the hockey rink, even though it’s a mile long, and you can score the point!

(b) Someone else will have to give the puck another hit along the way to make sure that it gets to you. A mile long is too far for the puck to go, even if it is a frictionless surface.

Explain your response: _____________________________________________

_______________________________________________________________

Figure 6: Sample application question from the evaluate portion of the inquiry cycle
Suppose that you are trying to get an ice hockey puck to travel along the track shown below (without hitting the sides). At the start of the track, somebody hits it in the direction shown.

Note that whenever anybody hits the puck, they always give it the same size hit.

(i) In which direction, A, B, or C, should somebody hit it so that the puck makes the first turn? (circle your choice): A B C

Figure 7: Sample two-dimensional motion Applied Physics Test question
The following choices are all definitions of the word ‘model.’ From the point of view of building a scientific theory, which is the best definition of a model? (Please circle one response)

(A) A small copy of an object

(B) A set of rules that allow you to predict and explain *

(C) A simplified or idealized picture of something *

(D) A set of plans for constructing a building or bridge

(E) A mannequin or someone who displays clothes

(F) I really don’t know!

(* The most sophisticated responses are in italics.)

Figure 8: Nature of modeling Model Assessment item
Suppose that there are two balls resting on a smooth floor, a striped ball and a white ball. They are the same size and shape, but the striped ball has four times the mass of the white ball.

Now suppose that you hit both balls at the same time, in the same direction. You give identical hits to both balls.

After you hit the balls, what can be said about their velocities?

(Circle your choice below)

(A) The striped ball will have the higher velocity.
(B) The white ball will have the higher velocity.
(C) Both balls will have the same velocity.

Figure 9: A sample Applied Physics Test item
APPENDIX A
Modeling Assessment

Name

Date:

Teacher:

Period:

ThinkerTools: Modeling Assessment

We want to know what you understand about scientific models and scientific modeling. Don’t worry if you don’t know exactly what a model is. Just do the best that you can in answering these questions. Your answers will be very important in studying the effectiveness of this curriculum.

If you don’t understand something, please ask the teacher. Don’t talk to other students, because we want to know what you think.
### SECTION B: Questions About Models

Question 7. Circle all of the items which you think are models:

<table>
<thead>
<tr>
<th>a scientific theory like Einstein’s theory of relativity</th>
<th>a pencil</th>
<th>a computer simulation like SIM CITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a bicycle</td>
<td>a globe or map</td>
<td>a snowflake</td>
</tr>
<tr>
<td>a rule like “roughly every twenty-four hours, the sun rises in the east and sets in the west because the earth rotates on its axis”</td>
<td>an orange</td>
<td>a video animation like the movie TOY STORY</td>
</tr>
<tr>
<td>an equation like Newton’s second law which says that force applied on an object is equal to the mass of that object times the object’s acceleration. ((F = m \times a))</td>
<td>a compact disc</td>
<td>a diagram of an atom</td>
</tr>
<tr>
<td>a set of diagrams and plans for a building or a room</td>
<td>a toy car</td>
<td>a picture or drawing</td>
</tr>
</tbody>
</table>

(Continue on the next page)
Question 8: Explain why the items you have circled are models.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Question 9: The following choices are all definitions of the word ‘model’. From the point of view of building a scientific theory, which is the best definition of a model?

(Please circle one response)

(A) A small copy of an object
(B) A set of rules that allow you to predict and explain
(C) A simplified or idealized picture of something
(E) A set of plans for constructing a building or bridge
(F) A mannequin or someone who displays clothes
(G) I really don’t know!

Explain your choice: __________________________________________________________

________________________________________________________________________
Question 10. If a scientist wanted to create a scientific model of an atom in order to predict how that atom will interact with other atoms, what parts of the atom would a scientist include in the model? (Please circle one response)

(A) every single part of the atom
(B) only the main parts of the atom
(C) only parts useful for predicting how it will interact with other atoms

Question 11: Can there be different kinds of models of the same thing? For example, are there different kinds of models for an atom? (Please circle one response)

(A) YES
(B) NO

Explain your reasoning using a different example: __________

Question 12: If you and your partner had different theories, do you think you could have a computer programmer create two different computer models for each of your theories? (Please circle one response)

(A) YES
(B) NO
If you think it’s possible to create two different computer models, explain how this might be useful. If you don’t think it is possible to create two different computer models, explain why it’s not possible.

Question 13: Could a scientist create an incorrect model?
(Please circle one response)

(A) YES
(B) NO

Question 14: Would a scientist ever change or revise a scientific model?
(Please circle one response)

(A) YES
(B) NO

Why or why not?

(Continue on the next page)
Question 15: Do you agree or disagree with the following statement? (Please circle one response)

“Even the best scientific theories and models aren’t necessarily true; they’re just ways of helping us understand the world.”

(A) Agree
(B) Disagree

Explain your choice: ____________________________
______________________________
______________________________

Question 16. Scientific models are: (Please circle one response)

(A) real and useful (models represent absolute reality)
(B) not necessarily real and useful (models don’t necessarily represent absolute reality)
(C) not real and not useful (models don’t represent absolute reality)

Explain your choice: ____________________________
______________________________
______________________________

(Continue on the next page)
SECTION C: Questions About Evaluating Models

Question 17: Do you agree or disagree with the following statement? (Please circle one response)

“Since scientists disagree about why dinosaurs became extinct, it’s clear that no one understands exactly how it happened. Therefore, any scientific model or theory of how it happened is just as good as any other.”

(A) Agree
(B) Disagree

Explain your choice: ______________________________________
_________________________________________________________
_________________________________________________________

(Continue on the next page)
Question 18: Do you agree or disagree with the following statement? (Please circle one response)

“When a scientist evaluates a scientific model, she looks for certain qualities such as how accurate and reasonable the model is.”

(A) Agree
(B) Disagree

If you agreed, describe some additional qualities. If you disagreed, explain why a scientist does not evaluate a model with certain qualities.

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Question 19: Do you agree or disagree with the following statement? (Please circle one response)

“Ways of evaluating scientific models or theories don’t change much over time ”

(A) Agree
(B) Disagree

Explain your choice: ________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
SECTION D: Questions About the Usefulness of Models

Question 20. From the scientific point of view, which is the best use of a model? (Please circle one response)

(A) to be a toy
(B) to copy an object or process
(C) to help someone construct an object
(D) to develop and test ideas

Question 21: Do you agree or disagree with the following statement? (Please circle one response)

“Computer models and simulations can help us understand things like the motion of a comet in space or traffic patterns in a city”

(A) Agree
(B) Disagree

Question 22. Two biology research groups have different models or theories about how a dangerous virus (like HIV) might replicate. How useful would it be for them to build and test computer models of each other’s theories?

(Please circle one response)

(A) Very useful
(B) Somewhat useful
(C) Not useful

(Continue on the next page)
Explain your choice: __________________________________________________________
____________________________________________________________________________
____________________________________________________________________________

Question 23. If you were an astronomer trying to determine the path of a comet in our solar system, which of the following would you rather have? (Please circle one response)

(A) A scale model (a smaller version of our solar system)

(B) A computer simulation or computer model of our solar system

Explain your choice: __________________________________________________________
____________________________________________________________________________
____________________________________________________________________________

Question 24: How could computer models help scientists with their research?
____________________________________________________________________________
____________________________________________________________________________
____________________________________________________________________________

Question 25: How could computer models help students learn science?
____________________________________________________________________________
____________________________________________________________________________
____________________________________________________________________________
Question 26:

Who do you agree with most? (Please circle one response)

(A) I agree almost entirely with Alexis
(B) I agree more with Alexis, but I think Dawn makes a good point
(C) I agree equally with Alexis and Dawn
(D) I agree more with Dawn, but I think Alexis makes a good point
(E) I agree almost entirely with Dawn

Why?

(Continue on the next page)
Question 27:

Who do you agree with most? (Please circle one response)

(A) I agree almost entirely with Justin
(B) I agree more with Justin, but I think Mark makes a good point
(C) I agree equally with Justin and Mark
(D) I agree more with Mark, but I think Justin makes a good point
(E) I agree almost entirely with Mark

Why?

________________________________________________________________________
________________________________________________________________________
SECTION D: Post-test Questions About the Computer Model

Question 28. You have conducted an experiment, analyzed your data, and come up with a rule. Then, you make a computer model that uses your rule. But, the computer model doesn’t behave like your real world experiment! How could this be?

Question 29: Suppose the computer is trying to figure out if the dot’s motion will change. Which of the following does the computer need to know?

(Please circle one response)

(A) How fast the dot is already moving
(B) The current position of the dot
(C) What forces (like friction and gravity) are acting

Question 30: How do you think the computer model works?

(Please circle one response)

(A) It uses rules like humans do to predict what will happen
(B) It reasons in a non-human machine-like way

Explain your answer:
Question 31. Suppose that a scientist is doing research and created a computer model which doesn’t behave like the real world. What could have caused this to happen?
FOOTNOTES

1 An enhanced multiple-choice question is a multiple-choice question with an additional short answer section that asks students to explain the reasoning behind their response.

2 A no-transfer problem involves concepts and contexts that have already been encountered in the curriculum. A near-transfer problem involves the same concepts but in slightly different contexts, a far-transfer problem involves the same concepts in entirely new contexts, and a beyond scope problem involves force-and-motion concepts that were not covered by the curriculum.

3 The data used from the prior ThinkerTools study are all seventh grade students from teacher A classes and classes from a teacher in another school.

4 We chose to use the scores from the entire test, as we wanted to compare these correlations with those from the prior ThinkerTools study.

5 This inquiry cycle is slightly modified from the one used in the prior ThinkerTools Inquiry Curriculum (see White & Frederiksen, 1998).

6 This question is similar to Grosslight et al.’s (1991) question, “Can a scientist have more than one model for the same thing?”

7 This question is similar to Grosslight et al.’s (1991) question, “Would a scientist ever change a model?”