COGNITION AND INSTRUCTION, *23*(2), 165–205

Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling

Christina V. Schwarz
Michigan State University

Barbara Y. White
University of California at Berkeley

We argue that learning about 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 (METT) 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. Curricular trials in four science classes of an urban middle school indicate that this approach can facilitate a significant improvement in students' understanding of modeling. Further analyses revealed that the approach was particularly successful in clarifying and broadening students' understanding of the nature and purpose of models. The METT Curriculum also led to significant improvements in inquiry skills and physics knowledge. Comparisons of METT students' performance with that of prior ThinkerTools students suggest that the acquisition of metamodeling knowledge contributed to these gains. In particular, METT students wrote significantly better conclusions on the inquiry test and performed better on some of the far-transfer problems on the physics test. Finally, correlational results, including significant correlations of pretest modeling and inquiry scores with posttest physics scores, suggests that developing knowledge of modeling and inquiry transfers to the learning of science content within such a curriculum. Taken together, the findings suggest that an emphasis on

---

Requests for reprints should be sent to Christina Schwarz, Michigan State University, 332 Erickson Hall, East Lansing, MI 48824. E-mail: cschwarz@msu.edu

model-based inquiry, accompanied by the development of metamodeling knowledge, can facilitate learning science content while also developing students' understanding of the scientific enterprise.

Models and the process of modeling are fundamental aspects of science. Advances in technology have emphasized their centrality by making model creation and revision even more accessible to scientists. Modeling tools provide new ways of creating theories, testing ideas, and analyzing data. Because students should understand the processes of science, and because modeling is fundamental to science and of great utility to scientists, many science educators have advocated model-centered instruction (Feurzeig, 1994; Feurzeig & Roberts, 1999; Lehrer & Schauble, 2000; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994). The National Science Education Standards emphasized this by stating that "all students should develop an understanding of the nature of science" and that this understanding includes knowledge that "scientists formulate and test their explanations of nature using observation, experiments, and theoretical and mathematical models" (National Research Council, 1996, p. 171).

Research indicates, however, that although students may successfully engage in creating scientific laws and models as part of inquiry-oriented science curricula, neither students nor their teachers typically possess much knowledge about the nature and purpose of scientific models (Carey & Smith, 1993; Grosslight, Unger, Jay, & Smith, 1991; Schwarz, 1998; Schwarz & White, 1998; Van Driel & Verloop, 1999; White & Schwarz, 1999). In this article, we refer to this kind of knowledge as *knowledge about modeling* or *metamodeling knowledge*.[1] We argue that without such metamodeling knowledge, students cannot fully understand the nature of science, and their ability to use and develop scientific models will be impeded.

Before further discussing the benefits and challenges of model-centered instruction and the importance of metamodeling knowledge, we clarify our use of the terms *scientific model* and the *process of scientific modeling*. For the purposes of this article, 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 such as $F = \mathrm{m}a$, to qualitative principles such as "when no forces are acting, an object's velocity remains the same, because there is nothing causing it to change." Models, in this sense of the term, are tools for expressing scientific theories in a form that can be

---

[1] We note that the term *metamodeling* has been used elsewhere (e.g., information systems development) with a different definition (the construct and rules needed for creating semantic models, see www.metamodel.com). We chose to use this term and define it as knowledge about modeling for the purposes of this article.

used for purposes like prediction and explanation. We use the term *scientific modeling* to mean the process used in much of modern science that involves (a) embodying key aspects of theory and data into a model—frequently a computer model, (b) evaluating that model using criteria such as accuracy and consistency, and (c) revising that model to accommodate new theoretical ideas or empirical findings. Our model-centered, metamodeling approach is one that places this modeling process, along with explicit instruction about the nature and purpose of models, at the center of the learning process.

BENEFITS OF A MODEL-CENTERED APPROACH WITH A FOCUS ON METAMODELING KNOWLEDGE

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 (J. Gilbert, 1995; Mellar et al., 1994; White, 1993a; 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 (diSessa, 1985; Papert, 1980; White, 1993a). Furthermore, modeling is increasingly important in society, and many students will need to use computer-modeling technology in their lifetimes (Sabelli, 1994; Tinker, 1993).

A model-centered, metamodeling approach, which emphasizes learning about the nature and purpose of models, 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 (S. Gilbert, 1991). 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; Songer & Linn, 1991). It is our contention that modeling curricula, which simply engage students in developing models, are not enough to achieve such epistemological sophistication (Carey & Smith, 1993). One needs to add a "metamodeling layer" to a modeling curriculum, which enables students to develop not only scientific models but also explicit theories about the nature of models themselves (Smith, Snir, & Grosslight, 1992; Wiser, Kipman, & Halkiadakis, 1988).

CHALLENGES OF A MODEL-CENTERED APPROACH

Even though model-centered instruction may accurately reflect 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 et al., 1991; Schwarz, 1998; Schwarz & White, 1998; White & Schwarz, 1999). For example, in a previous study (Schwarz, 1998), we found that students participating in an earlier version of our curriculum could not describe or explain the modeling phase of the inquiry cycle they were using, even though they frequently interacted with a computer model and were creating their own written laws or models.

Furthermore, 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. For example, Smith et al. (1992) and Wiser et al. (1988) developed physics curricula that included metaconceptual discussions about the nature of models. These curricula had no noticeable effect on students' level of thinking about models when pre- versus postinstructional interviews were compared.

OVERVIEW

In this article, we describe our efforts at addressing the challenges outlined earlier by developing an instructional approach that enables middle school students to learn about the nature of scientific models and the process of modeling, while they also develop inquiry skills and subject-matter expertise. To accomplish this, we created, taught, and evaluated the Model-Enhanced ThinkerTools (METT) curriculum in four seventh-grade urban classrooms for approximately 45 min a day for 10.5 weeks. Our approach provides students with (a) a language for talking about scientific modeling; (b) opportunities for carrying out experiments to test competing hypotheses; (c) software tools for creating, observing, and interacting with alternative simulation models that represent competing hypotheses about the laws of force and motion; and (d) opportunities for debating the merits of alternative models. We hypothesized that engaging students in the modeling process, along with discussion and reflection about this process for an extended period of time, would be effective in teaching students about scientific modeling. This article reports the effects of this instructional approach on student learning, and it characterizes the nature of the understanding that students derived from this approach.

Our research methods incorporate aspects of a design experiment (Brown, 1992; Collins, 1992). Our intention is to compare the METT curriculum with the prior ThinkerTools curriculum regarding students' learning of inquiry skills and physics content while assessing the nature and changes of METT students' understanding of modeling. The METT curriculum is similar to the original curriculum in many respects but has been changed to emphasize modeling. In the METT curriculum, after carrying out real-world investigations to test their hypotheses about force-and-motion phenomena, the students use the METT software to build simulation models that represent their ideas about the physical principles that produced the force-and-motion behaviors they observed. In contrast, in the prior ThinkerTools curriculum, the students used the behavior of the Newtonian simulation (White, 1993b), along with results from their real-world experiments, as the basis for formulating a set of rules to predict and explain force-and-motion behavior. In addition, both versions of the ThinkerTools instructional trials ran for similar lengths of time, with the same teacher and school. We therefore planned to test our hypotheses about the effects of metamodeling instruction by comparing outcome scores on our inquiry and physics assessments for the two versions of the curriculum. This comparison enables us to make inferences about the effects of teaching students about the nature of modeling on their inquiry skills and physics knowledge.

In conducting our research, we address three main research questions. First, "Does introducing an emphasis on model-based inquiry, accompanied by instruction in metamodeling knowledge, improve students' understanding of the nature and process of modeling?" Second, "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 conceptual physics knowledge?" Finally, "How is the development of modeling knowledge related to the development of science knowledge and inquiry skills?"

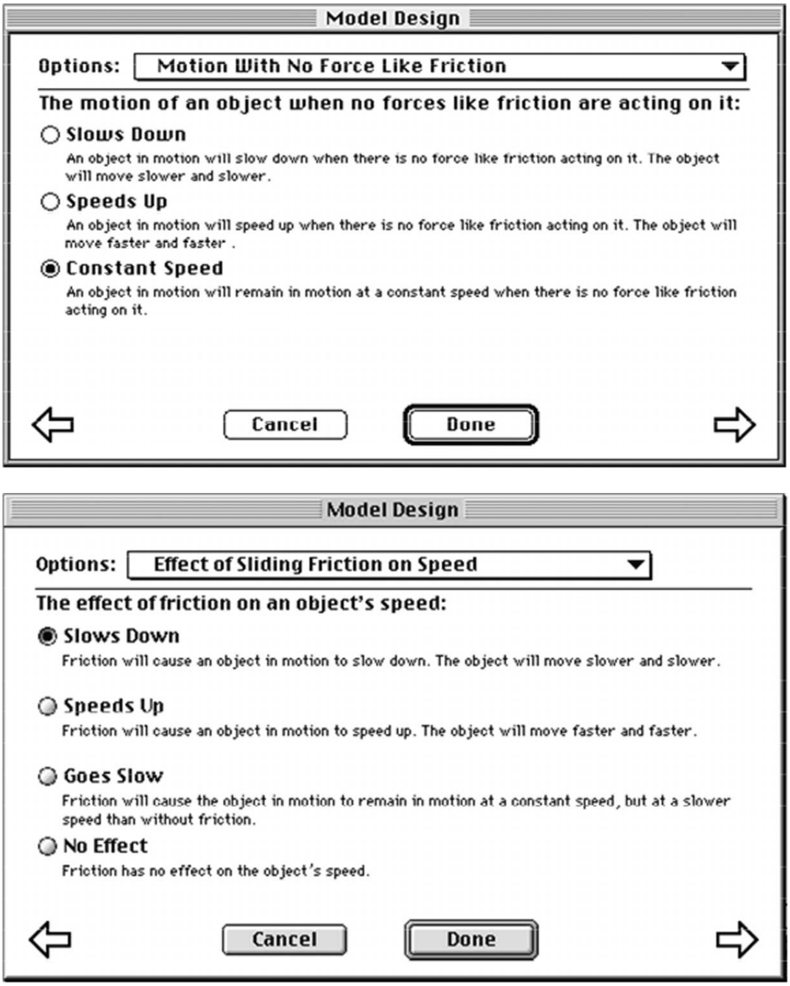
THE MODEL-DESIGN SOFTWARE

Before describing our METT curriculum, we first describe the Model-Design software, which is the primary modeling tool used in the curriculum. We, in collaboration with Christopher Schneider and John Frederiksen, designed the model-enhanced version of the ThinkerTools software (Schneider, 1998). 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 learn how to program. For a comparison of METT software with other modeling software, see our Web site.[2]

---

[2] The Web site can be found at http://www.msu.edu/~cschwarz

To enable students to investigate the implications of their own theories of force and motion, the METT software allows users to change the "laws" of motion that control the simulation software. Students do this by choosing from a set of alternative rules of force and motion, the rule that corresponds most closely to the physical law that they believe governs the real world. 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 1 for screen shots of two sets of rule choices offered by the software.)



Options: Motion With No Force Like Friction

The motion of an object when no forces like friction are acting on it:

○ Slows Down
An object in motion will slow down when there is no force like friction acting on it. The object will move slower and slower.

○ Speeds Up
An object in motion will speed up when there is no force like friction acting on it. The object will move faster and faster .

● Constant Speed
An object in motion will remain in motion at a constant speed when there is no force like friction acting on it.

Options: Effect of Sliding Friction on Speed

The effect of friction on an object's speed:

● Slows Down
Friction will cause an object in motion to slow down. The object will move slower and slower.

○ Speeds Up
Friction will cause an object in motion to speed up. The object will move faster and faster.

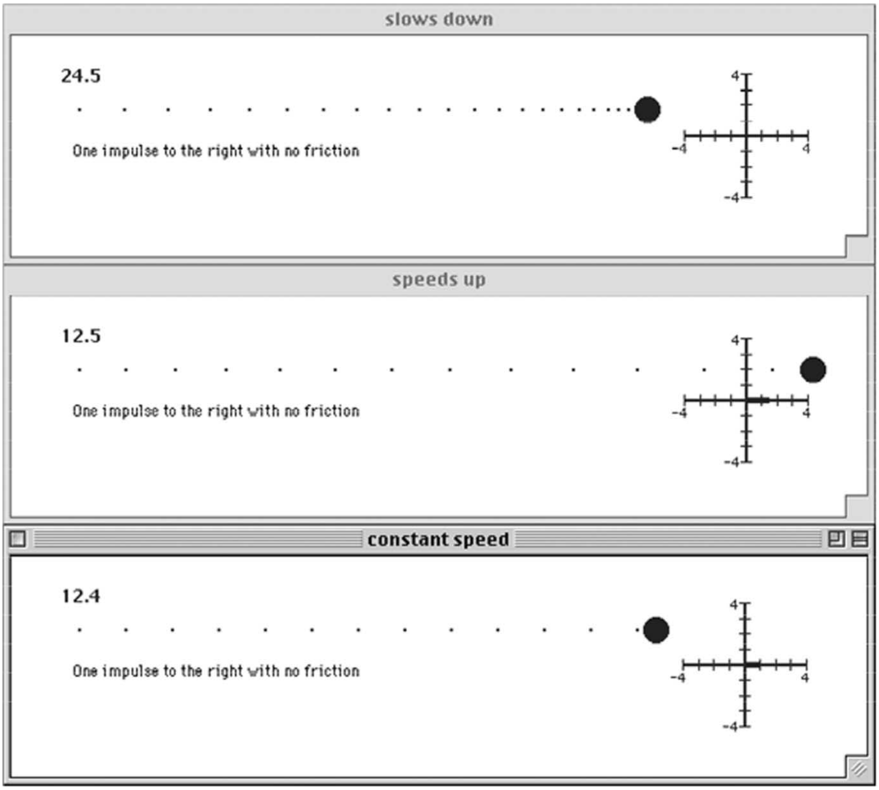
○ Goes Slow
Friction will cause the object in motion to remain in motion at a constant speed, but at a slower speed than without friction.

○ No Effect
Friction has no effect on the object's speed.

FIGURE 1 Screen shots illustrating how the software enables students to choose from among alternative laws of motion.

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 semiquantitative 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 to see the consequences of their chosen rule (see Figure 2).

As Figure 2 indicates, running the simulation 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



One impulse to the right with no friction

One impulse to the right with no friction

One impulse to the right with no friction

FIGURE 2 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 (small dots are left at fixed time intervals to provide a history of the object's motion, and thick lines on the "datacross" show the object's present velocity components).

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 employs, see White and Frederiksen (1998). Also, see Schwarz (1998) and Schneider (1998) for further discussion about the design and revision of this modeling aspect of the 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 (a) the nature of models (a model can be as simple as a rule that allows someone to predict a phenomenon; models are not necessarily real or correct, but good ones are better estimates of a phenomenon; there can be multiple models for the same phenomenon); (b) the nature of modeling (modeling involves embodying key parts of a theory into rules and representations); (c) the evaluation of models (models can be assessed using criteria like accuracy and plausibility); and (d) 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 as well as the other competing models. They were not told that the Newtonian model was the normative scientific model. Rather they were told that "Newton was a famous physicist from the 17th century who invented important models of force and motion." 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.

THE METT CURRICULUM

The METT curriculum is based on preliminary studies by Christina Schwarz (1996) and earlier work from the ThinkerTools Inquiry Project (White, 1984, 1993b; White & Frederiksen, 1998; White & Horwitz, 1988). The METT curriculum shares many features with the prior curriculum, known as the ThinkerTools Inquiry Curriculum, which focuses on teaching inquiry and the physics of force and motion (White & Frederiksen, 1998). Students in both the METT curricular approach and the prior curriculum conduct their research on force and motion by fol-

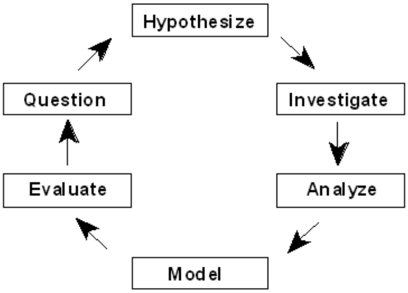


FIGURE 3 The ThinkerTools scientific inquiry cycle. This inquiry cycle is slightly modified from the one used in the prior ThinkerTools Inquiry Curriculum (see White & Frederiksen, 1998).

lowing the scientific inquiry cycle shown in Figure 3.[3] In the prior curriculum, they begin their inquiry by being presented with an interesting question about a situation involving force and motion, which is chosen to be simple (e.g., one-dimensional motion in a frictionless environment). They then develop hypotheses about what the motion of objects in that situation should be, based on their ideas about force and motion. Subsequently, they carry out investigations using real-world experiments and experiments with a computer simulation that follows rules of Newtonian physics. They then use their data to evaluate their hypotheses and summarize their findings by creating a law, which takes the form of a written rule that is consistent with their observations. Finally, they evaluate the limitations of their model (current set of laws) and consider situations and phenomena that it cannot explain. This cycle is then repeated as students expand their model of force and motion to encompass more complex situations, which include studying the effects of friction, motion in two-dimensions, and the effects of mass and gravity on the motion of objects.

The primary changes in the METT curriculum are the introduction of the Model Design software for students to use when they develop their model to account for the results of their real-world investigations instead of using the Newtonian simulation model as a source of data in their investigations. In addition, in METT, students receive explicit instruction about the nature of models and the process and utility of modeling and engage in discussions about these ideas. The text that follows describes a more detailed account of the METT implementation in the classroom.

---

[3] In the prior version of ThinkerTools, data analysis was incorporated within the "investigate" and "model" phases of the inquiry cycle. Further, the "investigate" phase was called "experiment," and the "evaluate" phase was called "apply."

THE METT CURRICULUM IMPLEMENTATION

At the beginning of the METT curriculum, the teacher (or in one class, Christina Schwarz) explained to the students that they would be learning about scientific modeling and that modeling is an important part of science. After completing a series of pretests, the students viewed and discussed a videotape showing modern uses of computer simulation models. The videotape included 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 from the video animation movie *Toy Story* (Lasseter, 1995). For an edited transcript that illustrates how some of the modeling language and concepts were introduced, please see our Web site. The students were then introduced to the curriculum and to the scientific inquiry cycle and science phenomena they would be investigating—the physics of force and motion.

To begin the inquiry cycle, students were given several “predictive questions,” and they formulated hypotheses about what happens to the motion of an object that is moving across a surface and why. For example, students are prompted with the following: “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?” An example of a student’s response is, “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 answer 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. They gave responses such as “it slows it down, but not as much.” They were then asked to form hypotheses about what would happen on a surface that is so smooth that there is absolutely no friction. They 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 about what happens and why.

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 applied an impulse (a standard-sized hit given with a mallet) to a smooth plastic puck and measured the object’s speed over various distances on a hard floor. 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. In this situation, the plastic puck was lifted from the surface it was running on by a cushion of air produced by an attached balloon.

This investigation 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 these "computer investigations," they would collect data about the behavior of the ThinkerTools Newtonian microworlds to help them test their hypotheses about competing principles that might govern force-and-motion phenomena (White, 1984, 1993b; White & Frederiksen, 1998; White & Horwitz, 1988). In contrast, students in METT used the software to represent and model their conceptual ideas, not as a source of evidence for testing their hypotheses.

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 formulate their models, based on the empirical evidence from their real-world investigations, and to help students reflect on the nature of models. In the beginning of this 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 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 that they observed in their experiments to see if they matched. The classes conducted this computer modeling activity and read passages about the nature of models and modeling. As we previously described, the students chose the computer-modeling rule that most closely corresponded to the rule that best fit their experimental data. Once students had chosen among the three (or four) computer rules, they ran the associated computer simulation to see the rule's consequences. They also ran the Newtonian computer model and compared it to the behavior of their model. For more details about students' interaction with the computer models, see a transcript of students working on their final project on our Web site.

In the first two modules, pairs of students also read and reflected on several reading passages about models and modeling. 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 to discuss advantages and disadvantages of different representations. Students read the passages and summarized the content to each other to reflect about the nature of modeling. For more details, see our Web site for a transcript of a conversation between two students about what defines a model.

Finally, at the end of the model phase of the inquiry cycle, after all students had finished both the computer modeling and reading activities, teachers engaged their students in a whole class discussion about the computer modeling rules and asked 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 written 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, 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.

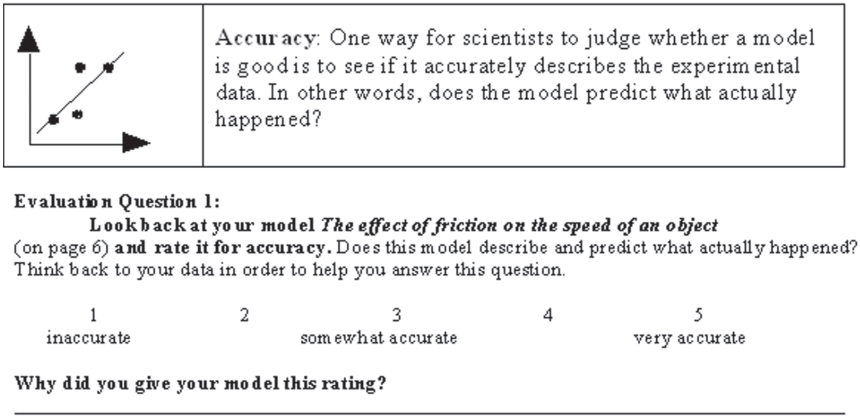
In evaluation, the last phase of the inquiry cycle, METT students evaluated their models with respect to modeling criteria such as accuracy and plausibility. This step was critical in helping students understand that models must be evaluated to establish their validity and to improve them. The evaluation activity was introduced in the first module by asking students 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 1 (*poorest fit*) and 5 (*best fit*) 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 4 from the research book.[4] 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 5 for accuracy and stated in her research book, "the real-world experiment and the computer model was [sic] about the same." Another student gave her model a 4 for plausible mechanism and stated,

---

[4] The research book is a student workbook that accompanied the curriculum. The entirety of the research book can be found in Schwarz (1998).



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.

1	2	3	4	5
inaccurate		somewhat accurate		very accurate

Why did you give your model this rating?

FIGURE 4 Sample model evaluation instructions.

"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 phase 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 [it stopped in mid air], 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 various situations. See Figure 5 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." Also, "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 were designed to help students apply and explore the implications of their conceptual models. As in prior ThinkerTools curricula, discussing such application questions leads students to see the limitations of their models, which helps them understand the need for including additional complexity in their subsequent models (White, 1993a). 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 (two-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 phase compared and contrasted their model to other stu-

Refer back to your ***Motion with no forces*** model and use it to answer the following question.

Application Question. 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 an answer and explain your choice in the space below.)

(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.

FIGURE 5 Sample application question from the evaluation portion of the inquiry cycle.

dents' models in class debates (cf. Bell & Linn, 2000; Koslowski, 1996; Kuhn, 1993). This debate activity was designed to motivate 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 1 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 evaluation phase of the prior version of the ThinkerTools curriculum differed significantly from the METT version. Students in the prior 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. They also engaged in a class discussion in which they tried to reach a consensus about the most accurate law. However, unlike the METT version, they did not engage in formal debates nor did they evaluate their models with the specific criteria of accuracy, plausibility, utility, and consistency.

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 fluid resistance. 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 curriculum were then given posttests, including a physics test, an inquiry test, and a scientific beliefs test. METT students also took a modeling knowledge test. We also note an additional key difference between versions of the curricula. Students in METT conducted less self-assessment than those in the prior ThinkerTools trial, and this self and peer assessment was shown to be a significant factor in students' learning of physics and inquiry (White & Frederiksen, 1998). Other differences included the number of modules in the curriculum (seven in the prior version compared to four in METT), and the number of topics from which students could choose in conducting their final research projects (nine in the prior version compared to three in METT).

EXPERIMENTAL DESIGN

This study addresses three primary questions. Does introducing an emphasis on model-based inquiry, accompanied by instruction in metamodeling knowledge, improve students' understanding of the nature and process of modeling? What effect does a metamodeling curriculum have on students' inquiry skills and conceptual physics knowledge? How is the development of modeling knowledge related to the development of physics knowledge and inquiry skills?

To investigate the knowledge of models and modeling, we developed a paper-and-pencil Modeling Test, which we administered before and after instruction in the METT curriculum, and we compared gain scores between pre- and postinstructional assessments. We also conducted postinstructional interviews with a sample of 12 students to obtain a more detailed picture of students' understandings of the nature of models and the purposes of constructing models. The interviews enabled us to determine the depth of students' modeling knowledge and the retention of modeling knowledge several months after the end of the curriculum. To address the effects of participating in the curriculum on students' understanding of inquiry and conceptual physics, we administered an Inquiry Test and a Physics Test before and after instruction. Because these tests were similar to those used in our prior study of the ThinkerTools Inquiry Curriculum (White & Frederiksen, 1998), this allowed us to compare students learning from the METT curriculum and with that from the prior curriculum. This allowed us to determine the impact that introducing modeling knowledge had on the development of inquiry skills and physics knowledge. Finally, to investigate how the development of modeling knowledge is related to the development of physics knowledge and inquiry skills, we carried out analyses of correlations among pre- and posttest scores on the Modeling, Physics, and Inquiry tests.

METHOD

Setting and Participants

A middle school science teacher and Christina Schwarz implemented the Model-Enhanced version of the ThinkerTools Inquiry Curriculum in four seventh-grade classes in an urban San Francisco Bay Area public middle school. The population of the school was ethnically, economically, and academically diverse. Approximately 44% of the school's students were Black, 31% were White, 13%

were Asian, 11% were Hispanic, and 1% of the students were composed of other groups. Additionally, 34% 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. 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 from the trial of the previous ThinkerTools curriculum.

Christina Schwarz cotaught one class of the curriculum with the teacher in this study, Ms. Jones. The teacher then used the curriculum without Christina Schwarz's presence in her remaining three classes. Ms. Jones was a 25-year veteran teacher who had taught the prior version of the ThinkerTools curriculum for 3 consecutive years.

Written Assessments and Analysis

While conducting this study, we used several different assessment instruments and data sources to address our research questions. These included paper-and-pencil pretests and posttests, student interviews, student work, classroom videotape, and Christina Schwarz's reflective journal. In this article, we focus on the findings from our written pre- and postassessments and from the interviews with students after the curriculum. There were three written paper-and-pencil tests: a modeling assessment, an inquiry test, and a conceptual physics test. Each written assessment was administered during one class period and did not affect students' grades in the class. In addition, Christina Schwarz conducted nature-of-models interviews with 12 students from one class 2.5 months after the curriculum and instruction ended.[5] The interviews provided in-depth insight regarding the students' understanding of modeling; they also indicated how well the students' understanding was retained over time.[6]

The instructional intervention and assessments were designed to address four aspects of modeling knowledge summarized in Table 1. We decided on these four aspects (nature, process, evaluation, and utility) by reviewing the literature (Carey & Smith, 1993; Grosslight et al., 1991) and then analytically determining key aspects of modeling knowledge. These aspects are not a definitive way of categoriz-

---

[5] We note that a complete analysis of data from an additional sample of two classes and 12 interviews with students from a different teacher exists on the Christina V. Schwarz's Web site (http://www.msu.edu/~cschwarz). Because the second teacher did not complete the curriculum, and to simplify the presentation for this article, we did not include those data.

[6] We note that in the intervening 2.5 months between the end of METT and the modeling interviews, the METT teachers taught a science sequence on spiders that was text-based and, to our knowledge, involved no modeling whatsoever.

TABLE 1
Types of Modeling Knowledge

- 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 or phenomena?
  - Constructed nature of models: Do models represent absolute reality?
- Nature or process 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?
- 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?
- Purpose or 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 or object?

ing modeling knowledge, but they served as a useful framework for creating the assessment and interview questions.

The Modeling Assessment utilizes a variety of question formats. These formats include a categorization 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?”).

The Scientific Inquiry Assessment, developed during our prior research (White & Frederiksen, 1998), gives students a research question and asks them to generate competing hypotheses, design an experiment, make up data and analyze it to reach a conclusion.

The Applied Physics Test incorporates items, designed to detect common force-and-motion misconceptions, which we have used to assess physics learning in our prior research (White, 1993b; White & Frederiksen, 1998).

Within our analysis of all pretests and posttests, we calculated students’ overall scores on the assessment, and 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 METT to results from the prior ThinkerTools Inquiry curriculum (in the case of the physics and inquiry assessments). Furthermore, we analyzed data from all of the written assessments using analyses of variance (ANOVA) that included four between-subjects factors. These factors included academic achievement level (low, ITAS score < 60; high, ITAS score > 60), prior ThinkerTools experience from the sixth grade (yes, no), gender (female, male), and Christina Schwarz's presence in the classroom (yes, no). Time of the test (pre and post) was also included as a repeated measures factor.

Modeling Interview and Analysis

The Modeling Interview included questions relating to the four aspects of modeling knowledge shown in Table 1 and lasted between 30 to 50 min. 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"). Christina Schwarz conducted the semistructured interviews, which took the form of a clinical interview in which students were occasionally asked to explain their responses or were reminded of additional evidence.

The 12 interview participants were students in one of Ms. Jones' classes. Ten students were girls, whereas 2 were boys; 9 students, or 75% of the interview sample, had ITAS scores higher than 60, the remaining 3 students had ITAS scores lower than 60. We interviewed all students from the class who volunteered and received parental permission to spend the time away from class.

Interviews were first transcribed from audiotape. From these transcriptions, Christina Schwarz summarized students' statements using a data reduction technique 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 to characterize and aggregate the students' responses to each interview question. Detailed information about how this coding scheme was developed and tested for reliability can be found in Schwarz (1998). Each student was given an overall level rating (strong, moderate, weak) for each interview question to characterize his or her overall response compared to responses that an expert might provide (such as those from Barbara White and her colleagues) as well as those from the literature such as Carey and Smith (1993) and Grosslight et al. (1991). (See Tables 3 and 4, shown later in this article, for indications as to how the overall level ratings were determined.)

RESULTS—MODELING KNOWLEDGE

In this section of the article, we turn to addressing our first 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 and evaluation, combined with instruction and reflection on the nature of scientific models and the process of modeling? To address this question, we determined the effect of METT on modeling knowledge by comparing pretest with posttest 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.

Overall Modeling Knowledge

We analyzed overall scores on the Modeling Assessment for four classes on the 18 multiple choice items of the assessment that are common to both the pretest and posttest. A summary of results from this analysis is given in Table 2. The posttest reliability of the total score on the Modeling Assessment, given by coefficient alpha is .81. The posttest reliabilities (coefficient alpha) for each of the four subscores are Nature of Models (.80), Process of Modeling (.20), Evaluation of Models (.19), and Purpose of Models (.47). Thus, students showed consistency in responses for items assessing their views of the Nature of Models and showed some consistency in items assessing the Purpose of Models, but no consistency for the other two categories of items. We should note that the items in the test were not subjected to any item analysis when the test was constructed.

Overall, there was a significant improvement in students' total score on the Modeling Test. Seventy-two students had an average pretest mean of 61% correct ($SD = 13$) and an average posttest mean of 70% correct ($SD = 14$), $t(71) = 6.48$, $p < .001$, $\sigma = .72$. In carrying out a repeated measures ANOVA, with time of testing (pretest and posttest) 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 was a significant effect of students' academic achievement level. With respect to academic achievement (as determined by scores on the ITAS assessment), students with lower academic achievement showed a mean gain of 3% compared to a mean gain of 11% for students with higher academic achievement, $F(1, 72) = 9.85$, $p = .002$.

To determine the relation between the Modeling Assessment and the Modeling Interview results, we correlated the total scores for these two assessments. We found a marginally significant correlation between the Modeling Interview scores and the Modeling Assessment posttest scores for the 12 students for whom we had both measures ($r = .53$, $p = .09$). This marginally significant correlation indicates that the written assessment is predictive of the modeling knowledge detected in the

TABLE 2
Overall Results of Student Model Understanding

Model Understanding	*Results From the Pre- to Postmodeling Assessment*	*Results From the Modeling Interview*
Overall results (including all aspects of modeling knowledge)	61% pretest, 70% posttest* for four seventh-grade classes on 18 items, $\sigma = .72$. Academic achievement a significant factor.	Evidence of sophistication in the various aspects. Evidence that knowledge was retained after the curriculum and instruction.
Nature of models	64% pretest, 73% posttest* on 7 items, $\sigma = .67$.	Evidence of sophisticated ideas about the nature of models, multiple models, and the constructed nature of models.
Nature of modeling or process of modeling	55% pretest, 58% posttest on 3 items.	Moderate sophistication about model revision.
Evaluation of models	54% pretest, 55% posttest on 3 items.	Mixed results. Students demonstrated strong use of model evaluation criteria, but were relativistic about "is any model as good as another?"
Purpose of models	57% pretest, 74% posttest* on 5 items, $\sigma = .55$.	Evidence of highly sophisticated ideas about the purpose of the model-design software and the general purpose of modeling.

$^*p < .001$.

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 Modeling Test as an assessment of modeling knowledge, because it may be useful for classroom teachers and researchers. It also suggests that the written assessment may be useful for supplementing richer assessment measures, like interviews, which must be conducted with smaller samples of students.

ANALYSES OF ASPECTS OF MODELING KNOWLEDGE

Aspect 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 phenomenon, and that models are constructed? We will begin with an analysis of some individual items from this aspect of the Modeling Assessment and subsequently report results from the modeling interviews. For the seven nature-of-models questions within the Modeling Assessment, the posttest reliability (coefficient alpha) was .80. Students scored an average of 64% in the pretest and 73% in the posttest, $t(71) = 5.62$, $p < .001$, $\sigma = .67$. For example, the first question asks students to identify which of 18 items they think are models. The items ranged from nonmodels, 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. The most dramatic results were changes in students' categorizations of the most abstract models, such as a causal rule, an equation, and a scientific theory. For instance, only 14% of students thought a causal rule was a model in the pretest, whereas 48% of them believed it was a model in the posttest, $\chi^2(1, N = 71) = 22.15$, $p < .001$. Similarly, only 21% of students in the pretest thought an equation like Newton's second law was a model compared to 55% of students in the posttest, $\chi^2(1, N = 71) = 22.15$, $p < .001$. Finally, only 21% of students thought a scientific theory was a model in the pretest, compared to 52% of students in the posttest, $\chi^2(1, N = 71) = 17.29$, $p < .001$. Of the six remaining items related to the nature of model aspect, one other item showed statistically significant change (pretest results indicated that students did not know how to define a model or defined a model as a small copy of an object whereas posttest results indicated that students moved toward defining a model a set of rules that allows you to predict and explain, $\chi^2(1, N = 61) = 4.17$, $p < .05$, and three other items related to multiple and incorrect models exhibited ceiling effects.

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 aspect of understanding about models related to students' definition of a model, the second to the idea that there can be multiple models of the same phenomena, and the third to the constructed nature of models.

Students' definitions of a model. Results from our analysis of students' responses to several questions about the definition of a model are given in Table 3. This analysis indicates that, several months after METT ended, interviewees showed considerable sophistication in their overall definitions of a model, replicating the findings from the Modeling Assessment. We found that 64% of students whose responses could be scored gave a strong definition of a model (that a model can be a predictive, explanatory rule, and it is not necessarily a physical object, but could be a mental object or made of words), and 27% gave a moderately strong definition. For example, one student stated, "[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."

TABLE 3

Frequencies of Interview Responses for the Definition of a Model

Codes for Definition of a Model	*Comment Codes*	*Level Rating*	*%*
A (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.		7	58
It's:			
1. a rule (that predicts and explains)	5		
2. a representation (that predicts and explains)	2		
3. a theory or idea (that predicts and explains)	2		
4. a reasoning structure (like a computer model which will fit your hypothesis or experiment) (that predicts and explains)	2		
5. a summarizing conclusion if it predicts and explains	1		
Properties (these do not count in the overall categorization):			
6. it doesn't necessarily have to be a physical object; it can be a mental object or words	1		
7. it should be accurate, plausible, useful, and consistent	2		
8. it is often simplified	0		
Bs (moderate): A moderate response is a two- or three-dimensional visual model.		3	25
It's:			
1. a diagram or design	2		
2. a picture or drawing	1		
3. a 3-dimensional object	1		
Properties (These do not count in the level ratings):			
4. it's useful	0		
5. you can see it visually	0		
6. it shows you something	2		
7. to show how things are made	1		
8. it helps construct objects	1		
Cs (weak): A weak response is one in which the student thinks of a model as a concrete object or something that is not a model.		1	8
1. a conclusion if it is a statement and not an abstraction that predicts or explains	1		
2. a smaller copy/version of an object	1		
3. to do it in the real world	1		
4. something scientific	2		
5. showing your work or what you learned	1		
6. a scientific gadget	1		
NA: Other/not clear; no information, not enough information, or interviewer did not ask the question.	1	1	8

Note. Interview questions were as follows: "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?" 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.

Students' understanding of multiple models. We analyzed several questions to determine student's knowledge about multiple models, which is another component of 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?"

Our analysis of the responses indicates that interviewees showed a fair amount of sophistication about multiple models. We found that 7 of 12 students (58%) had an average response that was strong, capturing the sense that different people may have different interpretive frameworks for the same data, whereas 1 (8%) had a moderate level response that captured the sense that different models arise because of aspects such as variability and error. Although 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.

Nature of models—constructed nature of models. We also explored students' understanding of the constructed nature of models, another component of this aspect 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?"

Our analysis of the responses indicates that interviewees showed some thoughtfulness and understanding about the constructed nature of models. Four students (33%) 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 stated, "[Models] don't tell exactly what will happen, but they'll tell us something that might wind up happening." Five students (42%) gave moderate and more relativistic responses, indicating that models are estimates because they are representations of the real world. Finally, one student (8%) gave a weak response, stating that she needed to personally see evidence to believe something was true. (Two other responses did not provide enough information to be scored.)

Aspect II: Nature or Process 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?

Analysis of Modeling Assessment

There are three items on the Modeling Assessment that address this aspect of modeling knowledge. The coefficient alpha for these items was .20, so no statistical analysis of the pre- versus posttotal score for the three items is warranted. One of these questions about the nature of modeling is the following: "If a scientist wanted to create a scientific model of an atom to predict how that atom will interact with other atoms, what parts of the atom would a scientist include in the model?" Forty-eight percent of students chose the best response, "only parts useful for predicting how it will interact with other atoms" in the pretest, compared to 46% in the posttest. The alternative responses included "every single part of the atom" for which 31% chose this in the pretest compared to 42% in the posttest and "only the main parts of the atom" for which 21% chose this in the pretest compared to 11% in the posttest. We conjecture that the increased response rate for "Every single part of the atom" in the posttest 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 provide evidence regarding students' understanding of the process of modeling. For example, 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 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 were designed to help students rethink the implications of their conceptual models, derived from their real-world experiments, and possibly revise their models.

Our analysis of the responses indicates that interviewees showed a moderate level of sophistication about revising models. We found that 4 of the 12 students (33%) had a response that was classified as strong, indicating that revision may occur by rethinking one's data and their implication as well as the purpose of the model. For example, one student stated, "Yeah, [scientists change or revise their models] because they might do the same thing [as me] and go home, … and be thinking about what they did in lab, … and be like, "yeah, that's a big mistake … 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." Six students (50%) had a moderate level response, capturing the sense that model revision occurs when there is new information or evidence, and two (17%) had a weak level response, indicating that model revision may not occur or would occur because the model was wrong. Students' degree of sophistication about model revision may have been limited because the curriculum emphasized model creation and reflection more than model revision.

Aspect 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?

Analysis of Modeling Assessment

There are three items on the Modeling test that address this aspect of modeling knowledge. The coefficient alpha for these items was .19, so no statistical analysis of the pre- versus posttest total score for these three items is warranted. For one of the items, there was a ceiling effect: Most students agreed in the pretest (92%) and posttest (93%) 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 pretest, 41% of all the students disagreed with this statement compared to 33% in the posttest. Neither change to these questions was statistically significant. In the case of the first item, however, this may be at least partly due to the ceiling effect. We need to create a larger set of 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. We report results from two questions from the interview that related to model evaluation—the first concerns 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 were also asked to evaluate a model that they had created.

Our analysis of students' responses indicates that they generally had a strong understanding and use of model evaluation criteria. Four of the 12 interviewees (33%) mentioned and used specific criteria to evaluate their models, obtaining a strong overall rating. For example, one student stated, "We [evaluated our model by] … looking at 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." Five students (42%) evaluated their model with respect to vague notions of correctness or clarity, receiving a moderate rating, and three other students (25%) 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." Although 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.

Our second research question for this aspect 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.

Our analysis of students' responses indicates that, although some students indicated a sophisticated understanding, many expressed a naive view of the value of alternative models. We found that 5 of 12 students (42%) gave a strong response stating that some models are better than others because of their validity or accuracy, whereas some models are odd or implausible. For example, one student stated, "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 … [where the] people were just kind of sloppy about it, then that model isn't as good as the other one." No students gave a moderate-level response that models have equal value when there is no way [for the student] to know which one is right. Six students (50%) gave a weak response, holding the relativistic notion that all models have equal value. Students whose responses fit in this category gave a variety of reasons to justify their response, including that people have worked hard on creating their models, people have different opinions, and there ultimately is no right answer. For example, one student stated, "Yes [any model is as good as another] because models aren't answers. They're just possibilities. … They're not facts … They're thoughts." The remaining student (8%) gave an unclear response.

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. The finding that student beliefs about model evaluation were inconsistent with their beliefs about the evaluation of models, combined with findings from the written assessment indicating the same inconsistency, may suggest that there are two "model evaluation" aspects rather than one.

Aspect IV: Purpose or Utility of Models

Finally, we assessed students' understanding of the purpose of models to ascertain whether students understood the purpose and utility of modeling.

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. Overall, on the five questions related to the purpose of models, students scored an average of 57% on the pretest and 74% on the posttest, $t(71) = 4.4$, $p < .001$, $\sigma = .55$. In one question about the purpose of models, for instance, students were asked, "From a scientific point of view, which is the best use of a model?" On the pretest, 41% of students chose the most sophisticated answer, "to develop and test ideas" compared to 60% in the posttest, $\chi^2(1, N = 63) = 7.2$, $p = .007$. Other possible responses to this question included, "to be a toy," "to copy an object or process," or "to help someone construct an object." On the four other questions related to this aspect, one other item showed statistically significant improvement (related to the benefits of computer simulation models over scale models) and two other items exhibited ceiling effects.

Analysis of Modeling Interview

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 each student was given an overall rating using a three-level rubric according to how well he or she understood the utility of modeling. Table 4 presents the coding scheme and frequencies of responses related to the purpose of modeling.

TABLE 4
Frequencies of Interview Responses for the General Purpose of Modeling

Codes for General Purpose of Modeling	*Comment Codes*	*Level Rating*	*%*
As (strong): A strong response indicates that models may be useful for envisioning, predicting, testing, or teaching.		10	83
1. to visualize one's own and other people's models (represent ideas)	6		
2. to test theories/make sure what you think is right (come to logical conclusions); compare models	5		
3. to manipulate models	1		
4. for predicting phenomena	5		
5. to test or investigate things not otherwise possible in the real world (including things too complicated to see in the real world)	4		
6. to help *explain* and show people what, why, or how something happens (about one's theories)	2		
7. to help figure out/*understand* what, why, or how something happens	4		
8. to help people create better models			
9. to teach people about the subject material of the model	5		
10. to teach students how to experiment, collect data and write out models			
Bs (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.		2	17
1. for constructing/making an object	4		
2. to get information	3		
3. to make things more accurate and easier than in real life; for convenience	2		
Cs (weak): A weak response indicates the student doesn't think that models can be useful.		0	0
1. they don't really help much because the computer just follows the ideas you put into it	1		
NA: Other/not clear; no information, not enough information, or interviewer did not ask the question	0	0	0

Note. Interview questions were as follows: "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?" 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.

Our analysis of students' responses demonstrates that most interviewees had a strong understanding of the purpose of scientific and computer models. We found that 10 students (83%) mentioned high-level purposes, such as visualizing one's own and other people's ideas, whereas two students (17%) discussed the purpose of models in a more vague way. One student with a strong response stated, "[Scientific models can be useful for] … predicting what you need to do . … We can see our own models and stuff. … So we can see stuff that can't really happen, that we can't really see [otherwise]."

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, 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 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 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 either less successful at promoting understanding about creating, evaluating, and revising models than we had expected, or that such understanding was not assessed well by the written modeling assessment. Furthermore, although 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, limitations in our assessments, or both.

RESULTS—INQUIRY AND PHYSICS KNOWLEDGE

We turn next to addressing our second major research question: How did the curriculum and instruction affect students' inquiry skills and conceptual physics knowledge? It is important to understand the potential benefits and drawbacks of an instructional approach that emphasizes metamodeling, particularly the effect on students' inquiry skills and science content knowledge, if we are to advocate and improve on this approach. 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 students' conceptual models of force and motion as well as their inquiry skills), we can determine how the different curricular emphases affected students' capabilities in these areas.

The two curricula differ in a number of respects, including the introduction of metamodeling instruction in METT, the use of only Newtonian simulations (as a source of data in the investigations) versus enabling students to construct other simulation models in METT, the number of modules and projects completed, and the use of self-assessment. This is a typical problem in design experiments, which is addressed by comparing the pattern of results found for the two versions of the curriculum and then considering competing interpretations for any differential effects that are found. In our case, to provide evidence for our hypotheses, we must show that the innovations introduced in the METT curriculum have strong and localized effects on particular aspects of students' inquiry skills and physics knowledge, which are likely to be attributable to the modeling emphasis in METT.

Inquiry Skills

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? We hypothesized that the METT approach would improve students' overall inquiry scores because it is an inquiry-based curriculum. We also hypothesized that the METT curriculum might improve students' performance in developing law-like models for their conclusions, compared to the prior curriculum, because METT students 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 students of the prior ThinkerTools curriculum.

Did students show evidence of improvement of their inquiry skills? If so, how do these results compare with those from the prior version of the curriculum?

The Scientific Inquiry Assessment

In assessing students' inquiry skills, we used the written Scientific Inquiry Assessment designed for instructional trials of the previous version of the ThinkerTools curriculum (White & Frederiksen, 1998). In this written assessment, students are asked to conduct an investigation that follows the inquiry cycle shown in Figure 3. The assessment begins by introducing students to a research question about the relationship between the mass of an object and the effect that sliding friction has on its motion (Does the weight of an object affect what friction does to its motion?). Students are then asked to develop and justify alternative hypotheses about possible answers to the question, design an experiment to test their hypotheses, create data they might obtain if they conducted their experiment, analyze those data, draw conclusions about those data, and then relate the conclusions back to their hypotheses. The instructions were requests to do each of these steps, such as, "Come up with two possible answers (hypotheses) to the question and give reasons for each."

Students' performance on the inquiry test was scored using an analytic method developed in the earlier ThinkerTools research (White & Frederiksen, 1998). This analytic method entailed coding 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 each hypothesis, and the quality of their explanations.

The inquiry tests were all scored blindly so that the scorer (Christina Schwarz) would not be influenced by the students' identity or whether the test was taken before or after the curriculum. A sample of Inquiry Assessments from two classes were selected for scoring using the criterion that each student who had completed the Inquiry Assessment also had to have completed all other METT pretests and posttests. This resulted in a sample of 38 sets of pre- to postinquiry tests. The reliability of the scoring was determined through double scoring 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.

Results and Discussion

Our analysis of students' performance on the Inquiry Assessment indicates that students showed a significant improvement in their inquiry skills due to participating in METT. We found that students had a mean of 47% on the pretest (*SD* = 21%) and a mean of 62% on the posttest (*SD* = 18%), $t(37) = 4.35$, $p < .001$, $\sigma = .71$. An analysis of variance, which included time of testing as a repeated measures factor and gender, standardized-test-score level, and previous ThinkerTools experience as between group factors, found no significant effects of the between group factors.

How do these results compare to those from the prior ThinkerTools study? Table 5 shows the data for Ms. Jones' students from both studies. There was no statistical difference between the inquiry posttest scores from the METT trial (62%) and those from her students in the prior ThinkerTools trial (56%), $t(83) = 1.41$, $p = .16$, for which posttests but no pretests were given to her students. 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.

Analysis of the subscores of the Inquiry Assessment (see Table 5) indicates no statistically significant differences between students' posttest subscores for METT versus the prior ThinkerTools curriculum, with the notable exception of the Conclusion Subscore. Students' Conclusion posttest scores from METT were significantly higher than those from the prior ThinkerTools trial, $t(84) = 2.43$, $p = .02$. The conclusions section of the Inquiry Assessment asks students to summarize the findings from their investigation. Their answers were scored as to whether or not they expressed their conclusions in the form of general law or model. Thus the improvement in scores on this section of the Inquiry Assessment lends support to our hypothesis that the METT approach should be more effective in enabling students to draw appropriate conclusions and form law-like models from their data. This may be due to their having a better idea of the form that a scientific model should take and of the criteria a good model should meet. Because this was one of the major goals of the METT curriculum, it is an encouraging finding.

Physics Knowledge

In this section of the article, 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 (how one creates and revises a model), as well as what how to evaluate models (what is a good model?), and the purpose of

TABLE 5
Scientific Inquiry Assessment Scores

	METT					
	Pretest		*Posttest*		*Prior TT Posttest*	
	M	*SD*	*M*	*SD*	*M*	*SD*
Total inquiry scores	47	21	62**	18	56	21
Hypothesis subscore	71	19	73	24	67	21
Experiment subscore	43	20	54**	15	51	18
Results subscore	30	23	62**	19	58	29
Conclusion subscore	38	32	59**	27	43	33
Coherence subscore	57	36	70*	29	66	33

Note. METT = Model-Enhanced ThinkerTools; TT = ThinkerTools. All data are represented as percentages.

$^*p < .05$. $^{**}p < .01$.

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 aspects of modeling 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.

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

The Applied Physics Test

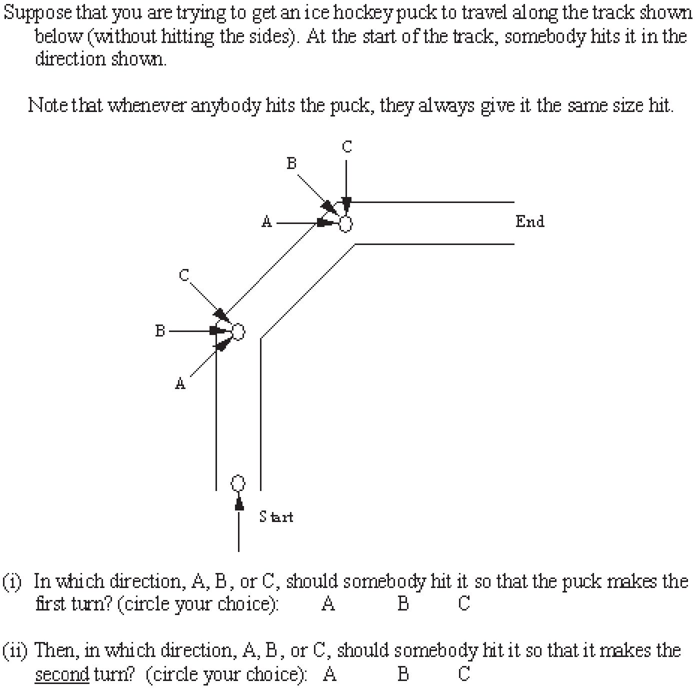
In determining students' physics knowledge, we used the Applied Physics Test, a paper-and-pencil assessment that was administered before and after the curriculum. The Applied Physics Test is similar to the one used in the prior version of the ThinkerTools curriculum and is comprised of 22 enhanced multiple choice questions[7] 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 (see the sample item in Figure 6). For further description of the assessment items, see White (1993b), White and Frederiksen (1998), and Schwarz (1998).

Overall Results

Analyses of 87 students from four METT classes indicates that students showed an improvement in their physics knowledge, with a mean of 49% correct on the pretest and 60% on the posttest, $t(86) = 4.9$, $p < .001$, $\sigma = .66$. The 12 items on this test included problems involving one- and two-dimensional motion and collisions in one dimension. 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 an 8.6% gain, with a pretest mean of 42% correct and a posttest mean of 51% correct, $t(85) = 3.14$, $p < .01$, $\sigma = .44$. The students from the prior version of the curriculum showed a gain of 9.6%, with a pretest mean of 49% and a posttest mean of 59%, $t(42) = 2.5$, $p = .02$, $\sigma = .50$. There is no statistical difference between gain scores from the two studies, $t(128) = .16$, $p = .87$.

---

[7] 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.



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

(ii) Then, in which direction, A, B, or C, should somebody hit it so that it makes the <u>second</u> turn? (circle your choice): A B C

FIGURE 6 Sample two-dimensional motion Applied Physics Test question.

By carrying out a repeated measures analysis of variance on the Applied Physics Test scores, we found that students' academic achievement level was a significant factor. Students with lower academic achievement on standardized test scores showed smaller gains on the physics test (*M* gain = 3%) than did students with higher academic achievement (*M* gain = 15%), $F(1, 83) = 7.11$, $p < .01$. Analysis of individual items, patterns in the data, and explanations of those patterns can be found in Schwarz (1998).

We note two findings of particular interest from that analysis. First, although METT improved students' responses on problems related to motion in two dimensions, their improvement was significantly less than improvement from students in the prior ThinkerTools. Complex model-design rules for this topic and less exposure to the Newtonian microworlds in METT may have made it difficult for students to develop Newtonian conceptual models that could be used to generate the correct answers to these problems. Second, METT students outperformed prior ThinkerTools students on a far transfer problem related to collisions in one dimension. We attribute this result to the modeling work within METT, in which students consider alternative computer models, and to the debates within the curriculum in which students argued about the nature of forces and their effects on the motion of objects.

CORRELATIONS AMONG MODELING, INQUIRY, AND PHYSICS EXPERTISE

In this section of the article, 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. To determine this, we calculated correlations between students' initial and acquired knowledge about modeling, inquiry, and physics using the written modeling, inquiry, and physics assessments. The sample includes 29 students whose inquiry pretests and posttests were analyzed, which represents all students from Ms. Jones's first two periods who had taken the pretests and posttests for all of the assessments. Correlations from a larger sample of students, which include data for students whose inquiry tests were not analyzed, show similar patterns and are reported elsewhere (Schwarz, 1998).

Several interesting findings can be seen in the pattern of correlations shown in Table 6. First, there are no significant correlations among the pretest measures of inquiry, modeling, and physics knowledge. The only appreciable positive pretest association is a correlation of .32 between scores on inquiry and modeling. As a result of participation in the METT curriculum, the correlation between inquiry and modeling knowledge increased, with a significant correlation between posttests of .50.

Although students' prior knowledge of inquiry and modeling are not correlated with their prior knowledge of physics, their pretest knowledge of inquiry and modeling is significantly correlated with the physics they learned from the METT curriculum (physics posttest scores), with correlations of .38 for the inquiry pretest and .47 for the modeling pretest. Further, the posttests of their inquiry and modeling knowledge developed within the curriculum also correlate significantly with their physics posttests, with correlations of .51 for the inquiry posttest and .47 for the modeling posttest. These patterns indicate that there may be a close link between knowledge of modeling and inquiry and the learning science content within

TABLE 6
Correlations of Pretests and Posttests

Learning Measures	*Modeling Pretest*	*Physics Pretest*	*Inquiry Pretest*	*Modeling Posttest*	*Physics Posttest*	*Inquiry Posttest*
Modeling pretest	1.00	—	—	—	—	—
Physics pretest	–0.24	1.00	—	—	—	—
Inquiry pretest	0.32	0.08	1.00	—	—	—
Modeling posttest	0.57**	–0.02	0.35	1.00	—	—
Physics posttest	0.47**	0.35	0.38*	0.47**	1.00	—
Inquiry posttest	0.37*	0.00	0.53**	0.50**	0.51**	1.00

Note. $N = 29$.
$^*p \leq .05$. $^{**}p \leq .01$.

the METT curriculum. This curriculum makes explicit to students how models of force and motion are developed over time through a series of investigations, and it engages them in this inquiry process as the basis for learning science content.

The overall pattern of results, including the significant correlations of pretest modeling and inquiry scores with posttest physics scores, suggests that developing knowledge of modeling and inquiry may transfer to the learning of science content in a model-based inquiry curriculum like METT.

SUMMARY OF FINDINGS AND IMPLICATIONS

Our findings indicate that a model-centered approach to science education, which emphasizes the development of metamodeling knowledge, can be effective in teaching students about scientific modeling, inquiry, and physics. The METT curriculum resulted in a significant improvement in students' performance on measures of modeling knowledge (a .72 σ gain) and inquiry skills (a .71 σ gain), which are notoriously difficult to teach, and a .67 sigma improvement in their performance on the physics test, for an area of physics known to be difficult for students (McDermott, 1984). The correlational evidence suggests that learning about the nature and purposes of models plays a role in the acquisition of inquiry skills and physics knowledge in the METT curriculum.

Analysis of students' performance on both the Modeling Assessment and the Modeling Interview indicates that our pedagogical approach succeeded in enabling many students, particularly students who scored above the median on a standardized achievement test, to learn about the nature and purpose of models. For instance, after completing the METT curriculum, many students were able to identify abstract representations as models and could show 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 curriculum came to understand 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 still needs refinement in promoting students' understanding of model evaluation and revision. After participating in the curriculum, many students still 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, although the students used model evaluation criteria during the curriculum, many gave posttest 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 equally valued, may have contributed to the lack of improvement on test questions about the comparative value of alternative models.

Our research raises important questions about the relation between metaknowledge and the use of knowledge in practice, as well as about the role of metaknowledge in learning. For instance, to what extent is it necessary for students to have explicit and sophisticated theories about the process of modeling to engage in model creation and revision (Schwarz, 2002)? In addition, how important is the goal of developing explicit knowledge about the nature of scientific modeling and inquiry, in addition to developing modeling and inquiry skills, compared to developing modeling and inquiry skills alone? Further, can these be made mutually supportive goals, as we tried to do in creating METT? The patterns of outcomes from the Modeling Assessment, the Scientific Inquiry Assessment, and the Applied Physics Test, as well as our correlational evidence, suggests that knowledge of physics, modeling, and inquiry become interrelated in the METT approach and may mutually support the development of one another. For example, we hypothesize that knowing about the various forms that models can take (Collins & Ferguson, 1993) may help students create models (such as those in the inquiry assessment), and understanding more about the utility of models may help students use them for scientific reasoning (as in doing the applied physics test). However, additional research is needed to further clarify this relation.

Given the potential benefits of learning about the nature and process of modeling, we propose that model-centered instruction with a focus on metamodeling knowledge should be incorporated more widely into science curricula, although further work is needed to develop the skills and knowledge in all students and within all areas of modeling knowledge. 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 10.5 week curriculum.

As we look toward the future, we see several promising areas for further research. In particular, it will be important to investigate authentic and motivating ways of incorporating modeling into curricula, as students often have difficulty understanding the purpose and payoff of modeling. In addition, learners will benefit from the development of innovative pedagogical approaches that combine model-exploration (enabling learners to explore previously created models such as the ThinkerTools Newtonian microworlds), with model-expression (enabling learners to create their own models within programming environments or in environments such as the Model-Design software in METT; Mellar et al., 1994). Finally, there is still a great need for the creation and experimentation of different kinds of computer-based modeling environments that enable students to work with a wide variety of models in multiple content areas. Research in these and other areas should lead to benefits for students and teachers alike, and will inform our understanding of scientific modeling and the role that modeling can play in the learning of science.

ACKNOWLEDGMENTS

This research was supported by a National Science Foundation science and design fellowship for Christina Schwarz and by grants to Barbara White from the National Science Foundation (Awards DPE–8400280 and REC–0087583) and the James S. McDonnell Foundation (Award 9135). We acknowledge the contributions of members of the ThinkerTools research team and other graduate students at University of California at 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 two anonymous reviewers, Annemarie Palincsar and Allan Collins, for their comments on earlier drafts of this article.

REFERENCES

Barowy, W., & Roberts, N. (1999). Modeling as inquiry activity in school science: What's the point? In W. Feurzeig & N. Roberts (Eds.), *Modeling and simulation in science and mathematics education* (pp. 197–225). New York: Springer-Verlag.

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*(8), 797–817.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2,* 141–178.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*(3), 235–251.

Collins, A. (1992). Towards a design science of education. In E. Scanlon & T. O'shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin, Germany: Springer.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28*(1), 25–42.

diSessa, A. A. (1985). Learning about knowing. In E. Klein (Ed.), *Children and computers, new directions for child development* (pp. 97–124). San Francisco: Jossey-Bass.

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

Feurzeig, W. (Ed.). (1994). Modeling and simulation [Special issue]. *Interactive Learning Environments, 4*(3).

Feurzeig, W., & Roberts, N. (Eds.). (1999). *Computer modeling and simulation in science education*. New York: Springer-Verlag.

Gilbert, J. (1995, April). *The role of models and modeling in some narratives of science education learning.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching, 28*(1), 73–79.

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching, 28*(9), 799–822.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: The MIT Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*(3), 319–337.

Lasseter, J. (Producer). (1995). *Toy story.* United States: Disney/Pixar.

Lehrer, R., & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology: Volume 5. Educational design and cognitive science* (pp. 101–159). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McDermott, L. (1984). Research on conceptual understanding in mechanics. *Physics Today, 37,* 24–32.

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

Mirtich, B., & Canney, J. (1994). *Impulse-based dynamic simulation* (Tech. Rep. No. 94–815). Berkeley: University of California at Berkeley, Computer Science Department.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas.* New York: Basic Books.

Sabelli, N. (1994). On using technology for understanding science. *Interactive Learning Environments, 4*(3), 195–198.

Schneider, C. (1998). *Facilitating model development through simulation: Formalizing student preconceptions into physical laws.* Unpublished manuscript, University of California, Berkeley.

Schwarz, C. (1996). *Student models, understandings and epistemological beliefs about computer models: A literature review and pilot study.* Unpublished manuscript, University of California, Berkeley, Graduate School of Education.

Schwarz, C. (1998). *Developing students' understanding of scientific modeling.* Unpublished doctoral dissertation, University of California, Berkeley.

Schwarz, C. (2002). Is there a connection? Perspectives on the role of meta-modeling knowledge in learning with models. In P. Bell, R. Stevens, & T. Satwicz (Eds.), *Keeping learning complex: The Proceedings of the Fifth International Conference of the Learning Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schwarz, C., & White, B. (1998, April). *Fostering middle school students' understanding of scientific modeling.* Paper presented at the annual meeting of American Educational Research Association, San Diego, CA.

Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction, 9,* 221–83.

Songer, N., & Linn, M. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching, 28*(9), 761–784.

Tinker, R. (1993). Modelling and theory building: Technology in support of students theorizing. In D. Ferguson (Ed.), *Advanced educational technologies for mathematics and science* (pp. 91–113). Berlin, Germany: Springer.

van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modeling in science. *International Journal for Science Education, 21*(11), 1141–1153.

White, B. (1984). Designing computer activities to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1,* 69–108.

White, B. (1993a). Intermediate causal models: A missing link for successful science education? In R. Glaser (Ed.), *Advances in instructional psychology, Volume 4* (pp. 177–252). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

White, B. (1993b). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1–100.

White, B., & Frederiksen, J. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 24,* 99–157.

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

White, B., & Horwitz, P. (1988). Computer microworlds and conceptual change: A new approach to science education. In P. Ramsden (Ed.), *Improving learning: New perspectives* (pp. 69–80). London: Kogan Page.

White, B., & Schwarz, C. (1999). Alternative approaches to using modeling and simulation tools for teaching science. In W. Feurzeig & N. Roberts (Eds.), *Computer modeling and simulation in science education* (pp. 226–256). New York: Springer-Verlag.

Wiser, M., Kipman, D., & Halkiadakis, L. (1988). *Can models foster conceptual change? The case of heat and temperature* (Tech. Rep. TR88–7). Cambridge, MA: Harvard Graduate School of Education, Educational Technology center.