Epistemologically Authentic Inquiry in Schools: A Theoretical Framework for Evaluating Inquiry Tasks

CLARK A. CHINN, BETINA A. MALHOTRA
Department of Educational Psychology, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

Received 1 January 2000; revised 8 December 2000; accepted 15 March 2001

ABSTRACT: A main goal of science education is to help students learn to reason scientifically. A main way to facilitate learning is to engage students in inquiry activities such as conducting experiments. This article presents a theoretical framework for evaluating inquiry tasks in terms of how similar they are to authentic science. The framework helps identify the respects in which these reasoning tasks are similar to and different from real scientific research. The framework is based on a recent theory of reasoning, *models-of-data theory*. We argue that inquiry tasks commonly used in schools evoke reasoning processes that are qualitatively different from the processes employed in real scientific inquiry. Moreover, school reasoning tasks appear to be based on an epistemology that differs from the epistemology of authentic science. Inquiry tasks developed by researchers have increasingly captured features of authentic science, but further improvement is still possible. We conclude with a discussion of the implications of our analysis for research, assessment, and instruction. © 2002 Wiley Periodicals, Inc. *Sci Ed* **86:**175–218, 2002; DOI 10.1002/sce.10001

INTRODUCTION

One of the central goals of science education is to promote scientific reasoning in students (AAAS, 1993; National Research Council, 1996). To this end, schools engage students in scientific inquiry tasks such as observation and experimentation. Even in curricula that are largely content oriented, hands-on inquiry activities play an important role. The goal of these activities is to provide a context in which students can learn to reason scientifically.

Our central argument in this article is that many scientific inquiry tasks given to students in schools do not reflect the core attributes of authentic scientific reasoning. The cognitive processes needed to succeed at many school tasks are often qualitatively different from the cognitive processes needed to engage in real scientific research. Indeed, the epistemology of many school inquiry tasks is *antithetical* to the epistemology of authentic science. If our argument is correct, our analysis has important implications for the design of school

Correspondence to: Clark A. Chinn; e-mail: cchinn@rci.rutgers.edu
Contract grant sponsor: National Science Foundation.
Contract grant number: 9875485.

reasoning tasks. New inquiry tasks will be needed that come closer to the epistemology and reasoning processes of authentic science.

This article is divided into two parts, as outlined below:

Part 1 presents a theoretical analysis of authentic scientific reasoning. We present our theoretical framework by contrasting *authentic scientific inquiry* with the *simple inquiry tasks* found in many textbook-based science curricula. We have chosen textbook inquiry tasks as our point of comparison because textbooks continue to be an important influence on science curricula (e.g., Driscoll et al., 1994; Kulm, Roseman, & Treistman, 1999; Stinner, 1995). In addition, the kinds of simple inquiry tasks found in science textbooks are prevalent in many other materials used in science instruction (e.g., Houghton Mifflin Interactive, 1997; VanCleave, 1997; Whalley, 1992). Therefore, we are contrasting authentic scientific inquiry with a form of inquiry that remains very common in science education. Our theoretical analysis of the features of authentic scientific inquiry (shown in comparison to simple inquiry tasks) provides the framework for the rest of the paper, as well as for the other four papers in this special section of the journal.

In Part 1, we begin by discussing two kinds of differences between authentic scientific inquiry and simple inquiry tasks. First, the *cognitive processes* needed to reason about simple inquiry tasks are often different from the cognitive processes used in authentic scientific inquiry. We develop a taxonomy of differences between cognitive processes employed in authentic science and the cognitive processes needed for simple inquiry tasks. Second, these differences in cognitive processes imply that the *epistemology* that underlies simple inquiry tasks is very different from the epistemology that guides authentic scientific reasoning. We present a second taxonomy analyzing differences between the epistemology of simple inquiry tasks and the epistemology of authentic science.

Next we present an explanation of why these differences in cognitive processes and epistemology exist. Applying Chinn and Brewer's *models-of-data theory* (Chinn & Brewer, 1996, in press; Chinn & Malhotra, 2001), we argue that any research study can be represented as a cognitive model. We show that the models that underlie authentic scientific research are qualitatively different from the models that underlie simple school inquiry tasks. We argue that these differences in models explain why reasoners must employ different cognitive processes and different epistemologies.

Part 2 of the paper applies the theoretical framework of Part 1 to provide a quantitative analysis of the characteristics of two groups of inquiry tasks: (a) inquiry tasks included in science textbooks, and (b) inquiry tasks that have recently been developed by researchers in the fields of education and psychology. The results of this analysis demonstrate a need to design improved inquiry tasks that incorporate more of the features of authentic science. We then discuss five general types of reasoning tasks in terms of their potential to capture various features of authentic science. We conclude by noting implications of our analyses for research, assessment, and instruction.

Part 1. Theoretical Framework for Analyzing Authentic Scientific Reasoning

Recent science standards have not only stressed the importance of learning to reason scientifically; they have also noted the complex nature of scientific reasoning. For instance, according to the *Benchmarks for Science Literacy* (AAAS, 1993, Sections 1a and 1b), students should learn that that there are many forms of scientific research, that observer bias is a threat to interpretation, and that different researchers using different methods can obtain different results. The National Science Education Standards (National Research Council, 1996; Olson & Loucks-Horsley, 2000) note that students should learn to develop theories that explain a diverse array of evidence, decide what evidence should be used, and critique explanations and procedures. These recommendations focus on helping students learn authentic scientific inquiry, not oversimplified forms of inquiry that are often found in schools (AAAS, 1993, p. 9). Thus, the standards highlight a need to develop a detailed, systematic analysis of the characteristics of authentic scientific reasoning. What are the core features of scientific reasoning? What cognitive and social strategies do scientists regularly employ when they engage in scientific inquiry?

The goal of this article is to begin the work of developing a systematic analysis of authentic scientific reasoning. Current analyses of scientific reasoning have not yet systematically specified key features of authentic scientific inquiry. The science standards (e.g., AAAS, 1993; National Research Council, 1996; see also Finley & Pocoví, 2000) point to important features of authentic inquiry, but they do not develop an analysis in detail. Many existing analyses of scientific reasoning focus on general categories of reasoning such as controlling variables, generating explanations, and providing evidence for explanations (Bybee, 2000; Germann, Haskins, & Auls, 1996; Hafner & Stewart, 1995; Kuhn et al., 1995; Zimmerman, 2000). These are all important features of scientific reasoning, but they are not unique to authentic inquiry. For instance, it is possible to control variables and provide evidence for explanations in very simple tasks that bear little resemblance to science. Other researchers have provided rich descriptions of interesting inquiry tasks but have not developed an extended analysis of the nature of authentic inquiry (e.g., Roth, 1995). This article aims to provide such an analysis.

Our analysis of authentic scientific inquiry is based on work in the psychology of science (e.g., Brewer & Mishra, 1998; Dunbar, 1995), the sociology of science (e.g., Kim, 1994; Latour & Woolgar, 1986; Pickering, 1995), the philosophy of science (e.g., Franklin, 1986; Galison, 1997; Giere, 1988; Kuhn, 1962), and the history of science (e.g., Rasmussen, 1993; Rudwick, 1985). Our analysis generally takes a cognitive approach. Cognitive analyses of scientific learning and reasoning have provided many important insights into the development of curricula and methods in science and mathematics (e.g., Bransford, Brown, & Cocking, 1999; Bruer, 1994). Cognitive analyses of science students' content knowledge, for example, have provided new insights into how students' alternative conceptions differ from scientific conceptions, and thus into how conceptual change topics can be taught more effectively (e.g., Driver et al., 1994). For example, cognitive analyses have provided insights into how to promote conceptual change (e.g., Driver et al., 1994; Guzzetti et al., 1993; Smith et al., 1997) and how to promote epistemological development (e.g., Carey et al., 1989; Smith et al., 2000). The work presented in this article follows in this tradition by employing cognitive analyses to gain insights into how to promote authentic inquiry in schools.

AUTHENTIC SCIENTIFIC INQUIRY VERSUS SIMPLE INQUIRY TASKS

Authentic scientific inquiry refers to the research that scientists actually carry out. Authentic scientific inquiry is a complex activity, employing expensive equipment, elaborate procedures and theories, highly specialized expertise, and advanced techniques for data analysis and modeling (Dunbar, 1995; Galison, 1997; Giere, 1988). Schools lack the time and resources to reproduce such research tasks. Instead, educators must necessarily develop simpler tasks that can be carried out within the limitations of space, time, money, and expertise that exist in the classroom. The goal is to develop relatively simple *school inquiry tasks* that, despite their simplicity, capture core components of scientific reasoning. Through carrying out these tasks, students are expected to learn to reason scientifically.

In this paper, we begin by drawing a contrast between authentic research and three prominent types of school inquiry tasks, which we collectively call *simple inquiry tasks*. Simple inquiry tasks appear regularly in textbooks (e.g., Daniel, Ortleb, & Biggs, 1995; McFadden & Yager, 1993), trade books (e.g., Murphy, 1991; VanCleave, 1997; Whalley, 1992), educational software (e.g., Houghton Mifflin Interactive, 1997; Theatrix Interactive, 1995), and websites of science activities (e.g., HIRO Science Lessons, n.d.; The Science House, n.d.), and so they are by no means uncommon in the science education landscape (see also AAAS, 1993). Our analysis will indicate that simple inquiry tasks incorporate few if any features of authentic scientific inquiry. Simple inquiry tasks are at one extreme of a continuum that ranges to authentic scientific inquiry as carried out by scientists. Of course, the curricula of many schools will fall somewhere in the middle of continuum, some closer to the authentic end, and others closer to the simple end. The important point is that our theoretical framework provides a systematic method for evaluating tasks to determine how closely they resemble authentic science.

As we analyze differences between authentic scientific research and simple inquiry tasks, it will be convenient to refer back to concrete examples of each type of reasoning task. Therefore, we begin by presenting two examples of authentic research and three examples of simple inquiry tasks.

Two Examples of Authentic Scientific Inquiry

Authentic scientific research takes many forms, from case studies in ecology to complex experiments using particle accelerators. Most scientific reasoning involves systematic comparisons of some kind, as in experiments, quasi-experiments, correlational studies, and comparative case studies. In this paper, we will illustrate some of the features of authentic research using two examples of real experiments: an early biological experiment investigating fermentation and a neuroscience experiment employing functional magnetic resonance imaging (fMRI).

Example 1: Study of fermentation. Buchner (1897/1955) conducted an early experiment investigating the process of fermentation. We will discuss two of the main conditions in this experiment. He began by grinding up and straining a mixture of Brewer's yeast, sand, water, and soil; this procedure yielded a strained liquid with no intact yeast cells. In one condition, he added glucose to the mixture; glucose was known to ferment in the presence of yeast. In a second condition, he added lactose, which does not ferment in the presence of yeast. Fermentation occurred only in the mixture with glucose, which showed that intact yeast cells were not necessary for fermentation to occur. Buchner proposed that fermentation is mediated by a chemical found in yeast that he called "zymase." Buchner further suggested that when intact yeast cells produce fermentation, the cells secrete zymase so that fermentation occurs outside rather than inside the yeast cells.

Example 2: fMRI study. Hirsch et al. (1993) used functional magnetic resonance imaging (fMRI) to investigate the effects of visual stimulation on neural activity, as indicated by increased oxygenated blood flow to specific regions of the brain. To provide an oversimplified overview, in fMRI studies a person lies motionless in a small space surrounded by a magnet that generates a powerful, uniform magnetic field. When placed in this magnetic field, paramagnetic atoms, especially hydrogen atoms, align their polarities with the field, effectively pointing them in the same direction. This alignment is then disturbed by introducing a radio wave frequency pulse. As the atoms return to their normal state, they emit signals during their decay that are measured by a detector. Because of differences in magnetic properties of oxygenated and deoxygenated blood, the decay rate in deoxygenated blood is greater than that of oxygenated blood. Through complex mathematical transformations, the decay signals are electronically converted into images in which higher densities of oxygenated blood in the brain are indicated by lighter pixels on an image.

The goal of the Hirsch et al. study was to investigate how visual stimulation affects patterns of blood flow in the brain. The researchers expected that visual stimulation would increase blood flow to three regions of the brain, called regions 17, 18, and 19. Participants were placed in a magnetic field that permitted four parallel cross sections of the brain to be imaged. Then a series of radio pulses was introduced. At each radio pulse, the researchers obtained images for each cross section of the brain.

The procedure was as follows. First, ten sets of brain images were made at 3-s intervals prior to any visual stimulation. Then each participant was shown a pattern of black bars on a white background, which was believed from prior research to have a strong stimulatory effect on visual regions of the brain. Ten more images were taken while the participant was looking at the pattern of bars. Next the pattern of bars was removed, and ten more images were taken while the participant was again without any visual stimulation. Images made during visual stimulation were compared statistically with images taken before and after stimulation, to try to determine which areas of the brain showed increased blood flow during visual stimulation.

Examples of Three Types of Simple Inquiry Tasks

In an analysis of the hands-on research activities in nine middle-school and upper-elementary-school textbooks (to be discussed in more detail later), we found that most hands-on research activities fell into three categories, which we call simple experiments, simple observations, and simple illustrations.

In *simple experiments*, students conduct a straightforward experiment, usually evaluating the effects of a single independent variable on a single dependent variable. For example, in one experiment in a middle school textbook (McFadden & Yager, 1993, p. 276), students affix a meter stick to the edge of a table so that the meter stick extends out from the table. Students then hang weights of various sizes to the end of the meter stick. The purpose is to investigate the effect of weight (the sole independent variable) on how far the meter stick bends (the sole dependent variable). When referring to this experiment later in the paper, we will call it the *meterstick* experiment.

In *simple observations*, students carefully observe and describe objects. In one typical exercise in Warner et al. (1991, p. 272), students observe a starfish, measuring features such as its diameter and noting the location of various structures such as the mouth and tube feet.

In *simple illustrations*, students follow a specified procedure, usually without a control condition, and observe the outcome. The experiment illustrates a theoretical principle, and the text clearly specifies what the theoretical principle is. For example, Thompson, McLaughlin, and Smith (1995, p. 315) presented an activity that we will call the *bleach task*. Students pour 20 ml of liquid laundry bleach into a large test tube and then add 0.5 g of cobalt chloride to the bleach. Students place their thumbs over the opening of the test tube to feel what happens (there is pressure from gas forming); then they insert a blown-out but still glowing match into the top of the tube. The textbook explains that the match ignites because oxygen is produced in a chemical reaction. Simple illustrations are inquiry tasks only in the narrowest sense. Students do encounter new empirical phenomena when they carry out the procedure, but they have no freedom to explore further.

DIFFERENCES IN COGNITIVE PROCESSES

In this section, we contrast the cognitive processes that are needed in authentic scientific inquiry with the cognitive processes that are needed in simple inquiry tasks. Table 1 summarizes key differences across the four types of research tasks: authentic inquiry, simple

TABLE 1
Cognitive Processes in Authentic Inquiry, Simple Experiments, Simple Observations, and Simple Illustrations

	Type of Reasoning Task			
Cognitive Process	Authentic Inquiry	Simple Experiments	Simple Observations	Simple Illustrations
Generating research questions	Scientists generate their own research questions.	Research question is provided to students.	Research question is provided to students.	Research question is provided to students.
Designing studies				
Selecting variables	Scientists select and even invent variables to investigate. There are *many* possible variables.	Students investigate one or two provided variables.	Students observe prescribed features.	Students employ provided variables.
Planning procedures	Scientists invent complex procedures to address questions of interest.	Students follow simple directions on how to implement a procedure.	Students follow simple directions on what to observe.	Students follow simple directions on how to implement a procedure.
	Scientists often devise analog models to address the research question.	Analog models are sometimes used, but students do not reflect on whether the models are appropriate.	Analog procedures are usually not used.	Analog models are sometimes used, but students do not reflect on whether the models are appropriate.
Controlling variables	Scientists often employ multiple controls.	There is a single control group.	Control of variables is not an issue.	Control of variables is usually not an issue.
	It can be difficult to determine what the controls should be or how to set them up.	Students are usually told what variables to control for and/or how to set up a controlled experiment.	Not applicable	Not applicable
Planning measures	Scientists typically incorporate multiple measures of independent, intermediate, and dependent variables.	Students are told what to measure, and it is usually a single outcome variable.	Students are told what to observe.	Students are told what to measure, and it is usually a single outcome variable.

Making observations	Scientists employ elaborate techniques to guard against observer bias.	Observer bias is not explicitly addressed, although measuring devices such as rulers are used.	Observer bias is not explicitly addressed, although measuring devices such as rulers are used.	Observer bias is not explicitly addressed, although measuring devices such as rulers are used.
Explaining results				
Transforming observations	Observations are often repeatedly transformed into other data formats.	Observations are seldom transformed into other data formats, except perhaps straightforward graphs.	Observations are seldom transformed into other data formats, except perhaps drawings.	Observations are seldom transformed into other data formats, except perhaps straightforward graphs.
Finding flaws	Scientists constantly question whether their own results and others' results are correct or artifacts of experimental flaws.	Flaws in experiments are seldom salient.	Flaws in experiments are seldom salient.	If students do not get the expected outcome, they often assume that they did the experiment incorrectly.
Indirect reasoning	Observations are related to research questions by complex chains of inference.	Observations are straightforwardly related to research questions.	Observations are straightforwardly related to research questions.	Observations are straightforwardly related to research questions.
	Observed variables are not identical to the theoretical variables of interest.	Observed variables are the variables of interest.	Observed variables are the variables of interest.	Observed variables differ from theoretical variables, but the text explains the link directly.
Generalizations	Scientists must judge whether to generalize to situations that are dissimilar in some respects from the experimental situation.	Students usually generalize only to exactly similar situations.	Students usually generalize only to exactly similar situations.	Students usually generalize only to exactly similar situations.
Types of reasoning	Scientists employ multiple forms of argument.	Students employ simple contrastive reasoning.	Students employ simple inductive reasoning.	Students employ simple deductive reasoning.

Continued

TABLE 1
Cognitive Processes in Authentic Inquiry, Simple Experiments, Simple Observations, and Simple Illustrations *(Continued)*

	Type of Reasoning Task			
Cognitive Process	Authentic Inquiry	Simple Experiments	Simple Observations	Simple Illustrations
Developing theories				
Level of theory	Scientists construct theories postulating mechanisms with unobservable entities.	Students usually uncover empirical regularities, not theoretical mechanisms.	Students uncover empirical regularities.	Students do experiments that illustrate theoretical mechanisms, but they do not develop or investigate theories.
Coordinating results from multiple studies	Scientists coordinate results from multiple studies.	Students do just a single experiment.	Students only make a certain range of observations at one time.	Students do just a single demonstration.
	Results from different studies may be partially conflicting, which requires use of strategies to resolve inconsistencies.	Not applicable	Not applicable	Not applicable
	There are different types of studies, including studies at the level of mechanism and studies at the level of observable regularities.	Not applicable	Not applicable	Not applicable
Studying research reports	Scientists study other scientists’ research reports for several purposes.	Students do not read research reports.	Students do not read research reports.	Students do not read research reports.

experiments, simple observations, and simple illustrations. In our analysis, we discuss six of the fundamental cognitive processes that scientists engage in when they conduct research: generating a research question, designing a study to address the research question, making observations, explaining results, developing theories, and studying others' research. The analysis is summarized in Table 1. Further details on cognitive processes in authentic reasoning can be found in Chinn and Malhotra (in press).

Although we will focus our analysis on scientific experimentation as one important type of scientific research, our analysis can be readily extended to other types of scientific research (see Chinn & Malhotra, in press; Chinn & Brewer, submitted). Thus, the framework presented in Table 1 can be viewed as an analysis of authentic scientific reasoning in general.

Generating Research Questions

In simple inquiry tasks, students are told what the research question is (e.g., find out what happens when you mix bleach and cobalt chloride). By contrast, in authentic research, scientists must develop and employ strategies to figure out for themselves what their research question is.

Designing Studies

We discuss several subprocesses involved in designing studies, as listed in Table 1.

Selecting Variables. In most simple inquiry tasks, students are told which of several variables to investigate, and the variables are usually perceptually salient, such as weight and the distance that a meterstick bends. In authentic research, scientists select their own variables from a very large pool of potential variables, and they often invent or construct variables that are conceptually embedded in the theories being tested. For instance, the variable "type of sugar" (glucose vs. lactose) is not a variable plucked directly out of the world; type of sugar is a theoretical concept that is embedded in theories of chemical composition.

Planning Procedures. Procedures in most simple inquiry tasks are straightforward, as students follow a short series of prescribed steps as in a recipe. In authentic research, procedures are complex and often require considerable ingenuity in their development. In the fMRI study, for example, scientists developed a complex system of procedures to present the stimuli in a controlled fashion, to introduce pulses, to capture images, and to analyze complex data.

In authentic research, scientists often construct procedures using model systems. For instance, Buchner employed an in vitro model, assuming that in vitro results would analogize to living systems. Using such analog models involves difficult decisions about whether the processes in the experiment overlaps sufficiently with the assumed processes in the real world to make the experiment meaningful. Analog models appear in simple inquiry tasks, but as we discuss in more detail later, students usually are not encouraged to reflect on whether the analog is appropriate.

Controlling Variables. In simple observations and simple illustrations, there are usually no control conditions. In simple experiments, what needs to be controlled is usually straightforward. For example, when conducting experiments to see whether seeds sprout faster in the light or the dark, students consider a few variables such as the type of seed used, the depth of the seed, the type of container, and the amount of water given. Once students understand the control-of-variables strategy, they can almost routinely go down a list of variables and make sure that all untested variables are held constant across the conditions.

In authentic research, by contrast, it can be very difficult to know which variables need to be controlled and how to implement proper controls. The reasoner needs a very good causal model of the processes being tested in order to know what to control. For example, in the fMRI study, there are a large number of variables that could potentially require control, such as the position of the participant, the presence of random thoughts that participants might have during the experiment, the intrusion of any extraneous visual images that a participant might glimpse, and the proper operation of the equipment. Even when deciding to control a variable, it can be difficult to decide what the control should be. What is the proper no-visual-stimulation control for looking at a pattern of black bars on a white background? Should subjects close their eyes so as to see nothing, or would closing eyes trigger other kinds of neural stimulation that could yield misleading results? What alternative control procedures exist? Such questions about control can be difficult to answer, and multiple control conditions are often needed.

In addition, scientists often employ special external controls to verify that procedures and equipment are operating as intended (Dunbar, 1999). For instance, when participants were being tested in the fMRI experiment, experimenters affixed a vial of cesium chloride to one temple and a vial of saline solution to another. These vials were captured in the cross-sectional images, and they served as a control to assess the amount of static in the pulse. Such external controls never appear in simple inquiry tasks. In short, controlling variables is much more difficult in authentic science than in simple varieties of school science. Scientists must build up a great deal of knowledge about the causal processes that operate under various conditions in order to determine what the proper controls are.

Planning Measures. In authentic experimentation, scientists measure many different variables, including measurements that serve as manipulation checks, measurements of intervening variables, and multiple outcome measures. In most simple experiments and simple illustrations, by contrast, there is just a single outcome measure, such as the number of centimeters that a meterstick bends.

Making Observations

In authentic research, scientists often employ special methods to guard against perceptual bias (Woodward, 1989). In the fMRI experiment, the observation process is automated through the use of electronic equipment, which enables the researchers to avoid perceptual error completely. Judging the intensity of the image at each pixel is also accomplished through an automated, computerized process. Such issues of guarding against perceptual bias seldom if ever arise in simple inquiry tasks. Students do use measuring tools such as rulers to make measurements more precise, but there appears to be little if any discussion of the issue of perceptual bias or of other techniques to avoid perceptual bias.

Explaining Results

Several important aspects of explaining results in authentic science are discussed below.

Transforming Observations. In the fMRI study, as in most scientific research, raw data undergo one or more rounds of data transformation (see also Latour & Woolgar, 1986; Lynch, 1988). The signals are transformed into numerical data about the intensity of the pixels. These data, in turn, undergo various transformations and analyses until they yield quantities that are submitted to statistical procedures. When generating their explanations, scientists do not try to explain the exact brightness of every pixel but rather the transformed data (see Woodward, 1989). In most simple inquiry tasks, by contrast, observations are straightforward, and there is no need for extensive data transformation. Raw observations are sometimes graphed, but even graphing is rare (Germann, Haskins, & Auls, 1996).

Finding Flaws. In authentic scientific research, methods are complex and uncertain, and scientists spend a great deal of time and effort worrying about possible errors in methods, both in their own work and in the work of others (Chinn & Brewer, 1993; Franklin, 1986). By contrast, simple inquiry tasks are so simple that there is little scope for finding flaws in methods. Relatively little can go wrong when hanging weights from metersticks. Ironically, simple inquiry tasks can lead students to become aware of experimental error but promote a very unscientific approach to responding to errors. When conducting simple inquiry tasks as part of science labs, students generally assume that if the results do not turn out right, they must have done the experiment wrong (Pickering & Monts, 1982). Thus, when students get unexpected results, they do not entertain the possibility that their hypothesis is wrong, and when they get expected results, they do not entertain the possibility that their procedures may be flawed. Students differ from scientists in both respects.

Indirect Reasoning. In authentic scientific research, the reasoning is often extremely indirect. In the fMRI study, conclusions about neural activity in the brain are inferentially linked through a complex chain of inferences to the brightness of pixels in an image. As a consequence, the variables that are manipulated and measured in real research (e.g., the presence or absence of a pattern of bars and the brightness of pixels) are not identical to the theoretical variables of interest (e.g., the degree of visual stimulation and the amount of oxygenated blood in a region of the brain). The manipulated and measured variables are connected to the theoretical variables of interest through these indirect chains of inference.

The reasoning in simple inquiry tasks is much more straightforward. In simple experiments and simple observations, the theoretical variables of interest are identical to the variables that the student manipulates. In the meterstick experiment, for example, the theoretical variables of interest are weight and the distance that the meterstick bends, and these are exactly the variables that are manipulated and measured. Simple illustrations, on the other hand, often do involve an indirect inference from observation to theory. For instance, in the bleach task students observe a match bursting into flame, which is taken to demonstrate the theoretical point that oxygen is present. The observation (a match bursting into flame) and the theoretical conclusion (oxygen is present) are different. However, students do not have to make this inference themselves because the text provides the inference directly. As a result, students do not have to worry, as scientists do, about whether the ignition of the match shows that *oxygen* is present, or that some other flammable gas is present, or even that the test tube has released tiny invisible flames that ignite the match.

Generalizations. As we have noted, scientists often work with model systems or samples that require difficult decisions under uncertain evidence about whether generalizations are possible. In simple inquiry tasks, generalizations are much more straightforward. In the meterstick experiment, for example, students are not asked by the textbook to discuss the extent to which this result generalizes to other situations.

Types of Reasoning. Simple inquiry tasks require only a limited range of reasoning strategies. Simple experiments require only a simple form of contrastive causal reasoning; for instance, if the meterstick bends more when more weights are hung, then one should conclude that increasing the weight makes the meterstick bend more. In sharp contrast, authentic reasoning requires the use of a broad array of diverse reasoning strategies. Examples include postulating unobservable mechanisms that could explain existing results, looking for flaws in experiments, finding ways to verify the validity of new methods, making indirect inferences, choosing between two or more theories that each have some explanatory successes, and devising indirect procedures to address questions of interest. Simple inquiry tasks leave out most of the reasoning processes that are characteristic of science.

Developing Theories

We discuss two aspects of developing a theory in authentic reasoning.

Level of Theory. In simple inquiry tasks, there is little concern with constructing underlying theory. Rather, the focus is on directly observable empirical phenomena (how weights affect the bending of metersticks or where the mouths of starfish are located). Authentic inquiry, by contrast, is directed at the development of theoretical mechanisms with entities that are not directly observable, such as molecules, enzymes, amino acids, magnetic fields, and polarized hydrogen atoms.

In simple illustrations, theoretical explanations sometimes play a role, but the text or teacher usually *presents* the theory (e.g., why the match ignites in the bleach task), so that students get no experience in constructing theoretical explanations on the basis of evidence. Indeed, simple illustrations do not provide *evidence* for a theory so much as they give the teacher an example to use when explaining the theory.

Coordinating Results from Multiple Studies. In simple inquiry tasks, students are seldom asked to perform multiple studies on the same topic. When scientists develop theories, they coordinate results from many different types of studies conducted at different levels of analysis. In biochemical research, some studies focus on the level of easily observable regularities (e.g., whether ground-up yeast cells can cause bubbles to form during fermentation), whereas others focus on underlying mechanisms (e.g., studies of exactly how zymase catalyzes biochemical reactions). Scientists develop interpretive strategies for coordinating results among these disparate studies. Moreover, because the results of these various studies sometimes conflict with each other, scientists also develop and employ heuristics for resolving inconsistencies in results.

Studying Research Reports

A prominent feature of scientists' research life is studying other scientists' research (Brewer & Mishra, 1998; Latour & Woolgar, 1986). Reading and hearing about other scientists' research plays a central role in all of the cognitive processes described above (see, e.g., Dunbar, 1995). Scientists read the literature to learn standard procedures for choosing experimental parameters such as the rate of centrifugation needed to separate out mitochondria or the magnetic settings needed for fMRI studies of blood flow in the brain. Other scientists' research helps inform researchers about what variables need to be controlled, what should be measured, how to devise new measures, and what kinds of conclusions will be considered acceptable in the research community. Scientists' conclusions are grounded in the theoretical and empirical work of other scientists. In real science the ratio of studying other scientists' research to conducting one's own research is relatively high.

By contrast, reading expert research reports plays almost no role at all in simple forms of school science. At most, students conduct their own research and make some reports to each other. But even then, students do not study a body of research that has passed review by experts in the field. In textbook science the ratio of studying others' research to conducting one's own research is low.

DIFFERENCES IN EPISTEMOLOGY

So far we have discussed differences between the cognitive processes employed in authentic scientific inquiry and the cognitive processes needed to complete simple inquiry tasks. We now turn to epistemology. Epistemology refers to people's basic beliefs about what knowledge is and when it should be changed. For example, a child might believe that scientific knowledge has a simple causal structure and that people should change their beliefs only when simple, obvious experiments can be conducted.

The cognitive differences between authentic science tasks and simple forms of school science tasks imply fundamental differences in epistemology. We think that simple inquiry tasks assume an epistemology that is *opposed* to the epistemology of authentic science. As a result, students who learn about scientific reasoning through simple inquiry tasks may actually learn a nonscientific epistemology.

In the following sections, we discuss several differences between the epistemology of simple inquiry tasks and the epistemology of authentic scientific inquiry. Table 2 summarizes these differences. We believe that this analysis applies generally to many forms of authentic scientific research, both experimental and nonexperimental (see also Chinn & Malhotra, 2001).

Purpose of Research

In real science, a central goal is to develop and refine theoretical models in response to evidence (Darden, 1991; Giere, 1988). The models generally employ unobservable theoretical constructs such as molecules, electron clouds, and forces. By contrast, the goal of most simple inquiry tasks is only to uncover easily observable regularities (e.g., plants grow faster in the light than in the dark) or the salient structure of objects (e.g., plants have stems and leaves), not to generate theories about underlying mechanisms. In short, the ultimate goal of most authentic research is the development and revision of theoretical models. The goal of most simple inquiry is a Baconian gathering of facts about the world.

Theory–Data Coordination

Authentic scientific research requires scientists to seek global consistency within a complex web of data and theories (Thagard, 1992, 1999). In simple inquiry tasks, students only have to seek local consistency between a conclusion and (usually) a simple study. Because simple tasks are generally straightforward, there is little need for complex theory–data coordination. As a result, science students are likely to develop an overly simple view of science, believing that science is a discipline that uses simple, reasoning patterns that are not applicable in other human endeavors, where data are messy. Simple inquiry tasks do not give students an opportunity to develop a more scientific epistemology in which one strives for global consistency between theories and data, even when data are uncertain and partially conflicting.

Theory–Ladenness of Methods

One of the fundamental epistemological features of authentic science is that methods are partly theory laden. We do not believe that methods are entirely theory laden. Scientists regularly do change their minds when they are convinced by methodologically sound studies

TABLE 2
Epistemology of Authentic Inquiry, Simple Experiments, Simple Observations, and Simple Illustrations

Dimension of Epistemology	Type of Reasoning Task			
	Authentic Inquiry	Simple Experiments	Simple Observations	Simple Illustrations
Purpose of research	Scientists aim to build and revise theoretical models with unobservable mechanisms.	Students aim to uncover a simple surface-level regularity.	Students aim to observe structures of objects.	Students aim to understand a provided theory.
Theory–data coordination	Scientists coordinate theoretical models with multiple sets of complex, partially conflicting data.	Students coordinate one set of observable results with conclusions about those observable results.	Students record what they see.	There is no theory–data coordination.
	Scientists seek global consistency.	Students seek at most local consistency.	Students seek at most local consistency.	There is no theory–data coordination.
Theory-ladenness of methods	Methods are partially theory laden.	Methods are not theory laden.	Methods are not theory laden.	Methods are not theory laden.
Responses to anomalous data	Scientists rationally and regularly discount anomalous data.	There is little scope for students to rationally discount data.	There is little scope for students to rationally discount data.	Data are rejected as erroneous results contradict expectations.
Nature of reasoning	Scientists employ heuristic, nonalgorithmic reasoning.	Students employ algorithmic reasoning to derive a conclusion from an experiment.	Students may employ various modes of reasoning about visual structures.	Students comprehend the provided explanation linking the theory to the data.
	Scientists employ multiple acceptable argument forms.	Students employ simple contrastive arguments.	Students often make no arguments.	Students make no arguments.
	Reasoning is uncertain.	Reasoning is certain.	Reasoning is certain.	Reasoning is certain.
Social construction of knowledge	Scientists construct knowledge in collaborative groups.	Students construct knowledge in collaborative groups.	Students construct knowledge in collaborative groups.	Students construct knowledge in collaborative groups.
	Scientists build on previous research by many scientists.	Students seldom build on any previous research.	Students seldom build on any previous research.	Students seldom build on any previous research.
	Institutional norms are established through expert review processes and exemplary models of research.	There are no institutional norm-setting processes.	There are no institutional norm-setting processes.	There are no institutional norm-setting processes.

that yield results contrary to their expectations (Chinn, 1998; Thagard, 1999). However, it is also true that methods and theories are sometimes entangled. Rasmussen (1993) has analyzed a debate in which cellular biologists who stained their samples before observing them through a microscope found evidence for a cellular structure known as mesosomes, whereas biologists using freeze fracture microscopy concluded that mesosomes did not exist. It was difficult to work out techniques that eventually led to near-consensus on this issue (for analogous examples, see Collins & Pinch, 1993; Latour & Woolgar, 1986). Concern about the reliability of methods leads scientists to develop heuristics to validate methods such as checking whether new methods can obtain results that were solidly established by old methods (Franklin, 1986; Hacking, 1983).

There is no such interdependence between theory and method in simple inquiry tasks. Methods are assumed to be reliable if directions are properly followed. Hence, students have no opportunity to develop an epistemology in which careful critical reflection on methods is important. Students will learn neither the importance of validating methods nor any specific techniques for doing so.

Responses to Anomalous Data

A closely related aspect of epistemology is how reasoners respond to data that are anomalous for their current theory. Because the set-up for simple inquiry tasks is so straightforward, if the students obtain anomalous data, the only rational response to the anomalous data is to change their hypothesis. For example, if some students were surprised to find that metersticks are flexible and bend more when greater weights are hung, they would have little rational grounds for discounting the surprising data.

In authentic scientific research, however, there are many different legitimate responses to anomalous data. Chinn and Brewer (1993, 1998) have identified eight possible responses to anomalous data that are made by scientists. One of the eight responses is to change one’s theory. The other seven responses involve discounting the data in some way, namely, by ignoring the data, rejecting the data, expressing uncertainty about the data, excluding the data from the domain of the current theory, holding the data in abeyance, reinterpreting the data, or making peripheral theory changes. A scientific epistemology acknowledges that all of these responses are highly rational in appropriate circumstances and embraces heuristics for judging when data should be accepted and when they should be discounted in one of these ways. Simple inquiry tasks fail to promote such an epistemology.

Nature of Reasoning

Simple inquiry tasks require the use of simple, often algorithmic strategies of reasoning. For example, reasoning in simple experiments can be captured by several simple, algorithmic rules: (a) If all variables are controlled except for factor X, and there is a difference in outcomes, then conclude that X is causal. (b) If all variables are controlled except for factor X, and there is no difference in outcomes, then conclude that X is not causal. (c) If the experiment is not properly controlled, then conclude that the results are indeterminate. All of these are straightforward inferences that a computer can easily be programmed to make. By contrast, reasoning in real science involves uncertain judgments and heuristics. Scientists may be unsure about every aspect of drawing inferences from experiments. Are the controls adequate? Are the differences large enough to draw any meaningful conclusion? What kinds of generalizations can be made with any confidence? How do these findings mesh with findings from other types of studies? Given that there are conflicting data, which model or theory is more believable? To answer these questions, scientists cannot use any easy-to-program rules but instead must use a wide range of fallible heuristics. This requires the use of many different forms of argument as scientists reach difficult judgments where hard-and-fast rules do not apply.

A consequence of inquiry tasks that invoke algorithmic reasoning is that students may come to see science as comprising *certain* knowledge derived from *simple* logical rules of reasoning. They will not learn that science is uncertain, constantly undergoing scrutiny and revision, employing heuristics that fall short of certainty.

Social Construction of Knowledge

Another feature of the epistemology of authentic science is the construction of scientific knowledge through social processes and institutions (Chinn, 1998; Knorr-Cetina, 1981; Latour & Woolgar, 1986; Thagard, 1999). School science captures some aspects of the social construction of scientific knowledge. In most school inquiry tasks, students work in groups to conduct and interpret scientific research, as scientists do. However, there other equally important aspects of the social construction of real scientific knowledge that are absent from simple forms of school inquiry. For instance, scientists build on each other's work in a way that is absent in simple school science. Scientists start with a firm grounding in the methods, theories, and empirical findings of science, which is acquired by studying other scientists' work. As we have noted, studying expert research is almost invariably absent from simple inquiry tasks. In addition, simple inquiry tasks typically lack certain institutionalized procedures found in science such as review of articles by experts. Such procedures help create institutional norms that provide general guidelines for scientists. Scientists are further aware that their field holds up certain papers as exemplary, as models for research methods and patterns of argument (Kuhn, 1962). In these ways, the social construction of knowledge proceeds in ways that go beyond simple collaboration in groups.

Summary

One important implication of our analysis is that simple inquiry tasks may not only fail to help students learn to reason scientifically; they may also foster a nonscientific epistemology in which scientific reasoning is viewed as simple, certain, algorithmic, and focused at a surface level of observation. Researchers have found that many students appear to hold such beliefs about science (e.g., Carey et al., 1989); our analysis suggests that simple inquiry tasks used in schools may be partly responsible for promoting these beliefs.

ACCOUNTING FOR THE DIFFERENCES IN COGNITIVE PROCESSES AND EPISTEMOLOGY: MODELS-OF-DATA THEORY

In the previous two sections, we have analyzed differences in cognitive processes and differences in epistemology between authentic scientific research and three common types of simple inquiry tasks. In this section, we present an explanation for why these differences exist, based on *models-of-data* theory (Chinn & Brewer, 1996, in press). Our explanation assumes that experiments and other forms of research can be represented as cognitive models. The cognitive models that underlie authentic experiments are fundamentally different from the cognitive models that underlie simple experiments, and the differences in models help account for why there are differences in cognitive processes and epistemology (see also Chinn & Brewer, in press; Chinn & Malhotra, 2001).

Here we briefly describe models-of-data theory, and then we show how the models underlying authentic experiments differ from models underlying simple experiments. We also briefly discuss how the differences in models give rise to the differences in cognitive processes and epistemology shown in Tables 1 and 2. Although we present examples of models of experiments in this paper, models of data can readily be constructed for nonexperimental research (see, e.g., Chinn & Brewer, 1996, in press; Chinn & Malhotra, 2001).

According to models-of-data theory, an experiment or other forms of research can be represented as a model that integrates theoretical explanations with the observations and with the details of the data gathering procedures. In an experiment, the overall model consists of a set of submodels, one for each condition in the experiment. The model can be diagrammed schematically as a semantic network consisting of events linked with four main types of connections: causal connections, inductive connections, analogical connections, and contrastive connections. In these diagrams (see Figures 1 and 2), each event appears as a block of text that summarizes what happened during that event. Causal connections between events are marked by arrows connecting two events. Sometimes two events jointly cause a third event; this is diagrammed as arrows from two events converging on the third event. As Figures 1 and 2 show, most of the events in models of experiments are connected to each other in causal paths. Contrastive connections are indicated by a dotted line between contrasted sets of events. Inductive generalizations from one set of events to another are symbolized as a single line, and analogical generalizations from one set of events to another are denoted as a double line. Inductive generalizations generalize from a subgroup to the whole group (e.g., generalizing from a sample of 100 patients who got better when taking a drug to all patients similar to these 100). Analogical generalizations generalize from one situation to a qualitatively different situation (e.g., generalizing from a test tube reaction to a reaction in living cells).

As illustrations of models of data, we present models of each of the authentic experiments that we described earlier. Figure 1 displays a model of Buchner's fermentation experiment. The model represents the understanding of a person who has read an article describing Buchner's experiment. The model shows two of the conditions in the study. The two conditions are represented by separate "submodels" that are separated by a dotted line denoting the contrast between the two experimental conditions. The submodel for each condition shows the sequence of events that is believed to occur in that condition. Some events are part of the experimental procedure; these events are marked with an asterisk in the figure. Other events are hypothesized to occur as a consequence of the procedures; these events have no asterisk. The parts of the contrasting submodels that differ are highlighted in Figure 1 as underlined text within the events.

Most of the events in each condition fall within a path of causally related events. Each condition consists of a causal path that traces the steps in the procedure (e.g., mixing the initial ingredients; pressurizing and grinding the mixture) and the hypothesized consequences of these steps (e.g., a strained liquid with zymase but no intact yeast cells is formed; bubbles form in the glucose condition). Some of the events are formally observed and recorded (e.g., gas bubbles appear inside the container); these events are enclosed in a rectangle in the diagrams. Other events are not directly observed or cannot be directly observed, because they involve unobservable theoretical constructs (e.g., CO_2 is evolved; zymase catalyzes the fermentation reaction).

As we noted earlier, Buchner made additional speculations about the fermentation process when intact cells are involved. Buchner proposed that fermentation probably occurs outside the cell as the cell secretes zymase outside the cell. This is a speculation that goes considerably beyond the available data, and it represents an analogical induction from the experimental context to the different context of fermentation with intact yeast cells. The site of the analogical induction is represented in Figure 1 as a double line.

As a second example of a model of an experiment, a partial, simplified model of the fMRI experiment discussed earlier is presented in Figure 2. This partial model shows the submodel

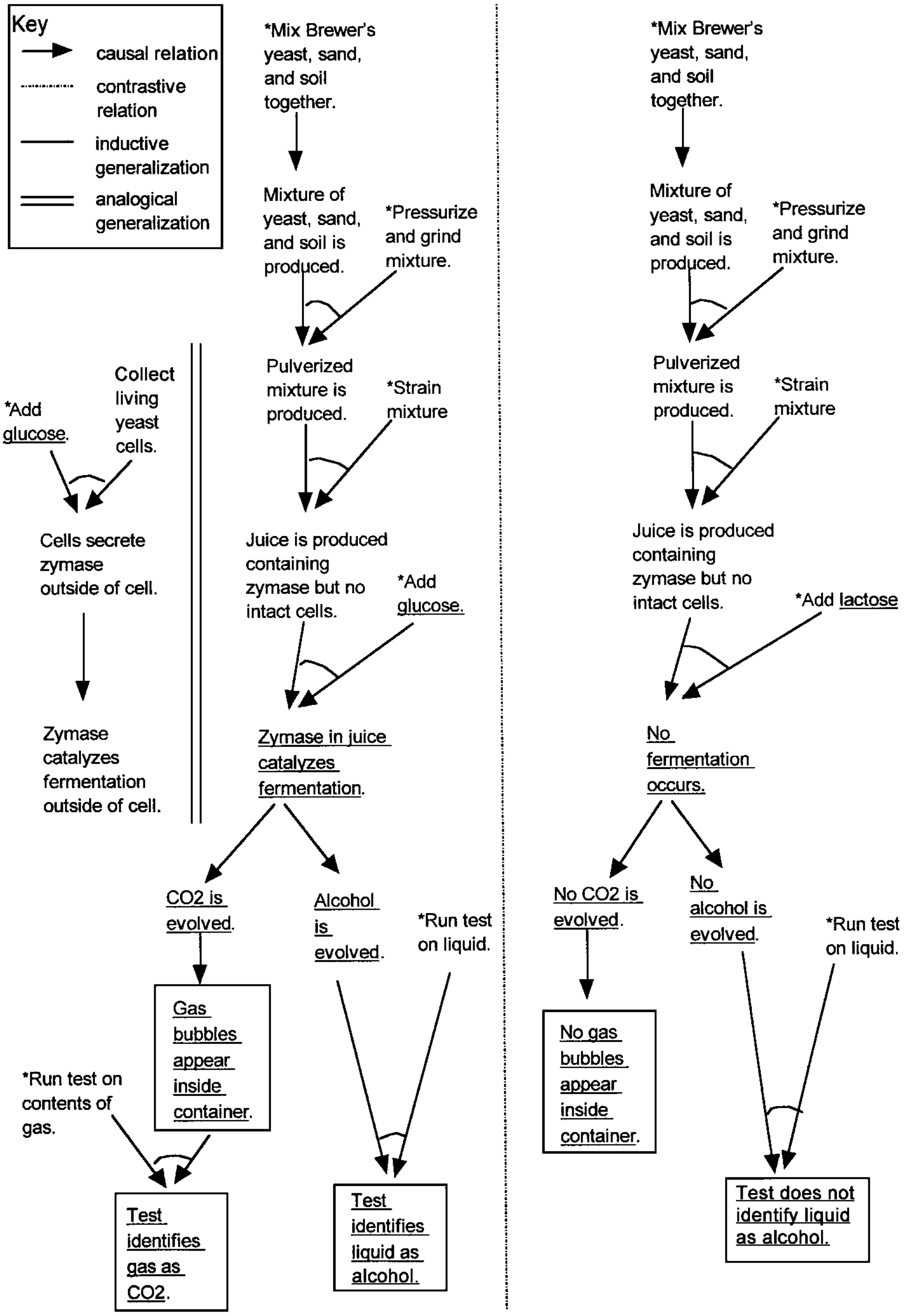

Figure 1. Model of Buchner's fermentation experiment.

for just one of the conditions in the experiment, the condition with visual stimulation. In the complete model of this experiment, there would be an additional submodel for other experimental conditions, including the condition in which no visual stimulation occurs. The events in the visual-stimulation condition that are expected to differ from the no-visual-stimulation condition are highlighted as underlined text in the figure.

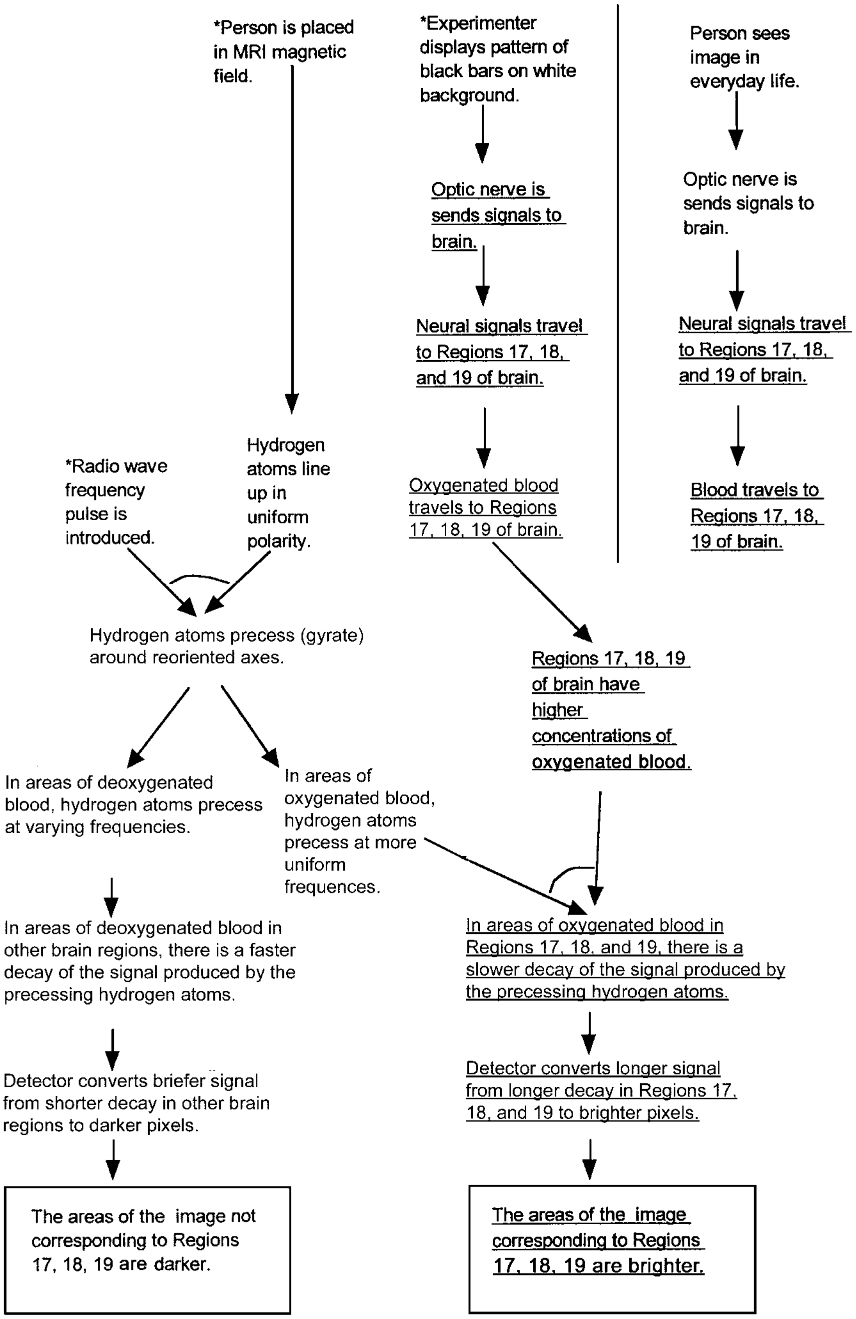

Figure 2. Submodel of one condition of the fMRI experiment.

Like Figure 1, Figure 2 shows a complex causal web with many intervening theoretical events (such as the theorized alignment of hydrogen atoms with the magnetic field). The model consists largely of interlinked causal paths. One part of the causal path traces the process by which a visual image produces greater oxygenated blood flow to regions 17, 18, and 19 of the brain. Another part of the causal path traces the effects of the magnetic pulse. This part of the model specifies the processes by which regions of oxygenated and deoxygenated blood are translated into areas of brighter and darker pixels in the image. Overall, the causal web traces the processess by which neural activity is eventually translated into lighter pixels on the fMRI images.

There is an analogical generalization made across the line in Figure 2. The model assumes that the neural processes that occur when an individual observes a pattern of bars while lying in an MRI machine generalizes to the very different situation of people in everyday life observing ordinary visual stimulation.

The models of data shown in Figures 1 and 2 integrate hypothesized theoretical processes with the procedures and observations in the experiment. We view models of data as cognitive entities, existing in the minds of those who conduct or learn about the experiment. Figures 1 and 2 show cognitive models that might be constructed by nonscientists who have a reasonably good lay understanding of the experiments. Scientists working directly in these fields would undoubtedly have more elaborated models.

In contrast to the relatively complex models of Figures 1 and 2, models of simple experiments have very few nodes and links. Figure 3 presents a model of two conditions in the meterstick experiment. Each condition consists simply of two events: the hanging of a weight and the bending of the meterstick, which is observed by the student. There are no other events assumed to be operating. The entire model consists of two contrasting conditions, each with just two nodes connected by one causal link.

An analysis of the differences between the models of authentic experiments in Figures 1 and 2 and the model of the simple experiment in Figure 3 helps explain the differences in cognitive processes and epistemology between authentic inquiry and simple inquiry tasks. Here we highlight several key feature of models of authentic experiments that are absent from models of simple inquiry tasks. (Although we will discuss authentic experiments and simple experiments, a very similar analysis holds for other types of research.)

Intervening Events in Models of Authentic Experiments

An obvious difference between models of authentic experiments and models of simple experiments is that models of authentic experiments have many intervening events between the initial and final events in each condition, whereas models of simple experiments have only the initial and final events. The intervening events in models of authentic experiments are important for several reasons. First, they often specify the theoretical mechanisms (such as zymase in Figure 1) that are hypothesized to mediate the change from theoretical mechanisms to observed events. Second, the intervening events provide multiple sites for potential measurements. Each event in a model can become the focus of a separate measurement. For instance, in the experiment diagrammed in Figure 1, the scientist could observe the juice

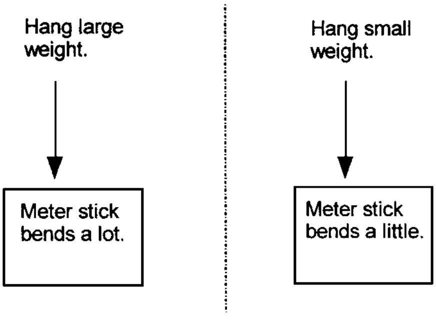

Figure 3. Model of a simple experiment.

through a microscope to directly verify that there were no intact cells, or the scientist could try to devise ways to indirectly measure the presence of the postulated enzyme.

As a related point about measurement, models of authentic experiments include sites where causal paths branch out. For instance, in Figure 1 the causal model for the fermentation process branches into two separate causal paths, one involving the production of carbon dioxide which bubbling within the solution, the other involving the production of alcohol that remains a liquid. Whenever a model of data branches out in this way, each branch can become the site of a separate test or measurement. In Figure 1, the scientist tests each of the chemical products to verify their identities. In this way, all of the intervening and branching outcome events in models of authentic experiments provide sites for potential measures. Because models of simple experiments lack both intervening events and branching outcome events, these models do not afford multiple measures.

Overall Greater Complexity of Models of Authentic Experiments

Models of authentic experiments are obviously more complex than the two-event models of simple experiments. There are not only intervening events and branching outcomes, as noted above; there are also complex causal branches at which further human intervention occurs. Because of the complexity of the model, designing an experiment becomes a matter of creativity and ingenuity. Experimenters must work out which events will be included in the experiment and how they will be arranged, which events to control for and how to control them, which events to measure and how to measure them. Each aspect of the construction process—selecting variables, planning procedures, planning controls, and planning measures—involves decisions that are made difficult because of the complexity of the model. In contrast, the models of simple experiments are so simple that designing experiments is trivial.

The complexity of models of authentic experiments also explains why there are so many ways in which an authentic experiment can be criticized. A model of an experiment can be criticized by finding an alternative cause for any one of the many events in the model (see Chinn & Brewer, in press). For example, in the fermentation experiment of Figure 1, a critic might argue that the CO_2 evolved not because of zymase-based fermentation but because of an entirely different chemical process. Because an experiment can seldom rule out alternative causes for all events in a complex model, reasoning about an authentic experiment cannot be algorithmic. By contrast, finding fault with a simple experiment is a straightforward processing of checking whether a few relatively obvious variables are controlled at the start of the experiment.

Analogical and Inductive Links in Models of Authentic Experiments

Models of simple experiments have only two kinds of links: a single causal link in each condition and a contrastive link separating the two conditions. Models of authentic experiments have analogical and inductive connections, as well. These additional connections inject a great deal more uncertainty into the judgments needed to evaluate authentic experiments. There is no way to be sure that an inductive or analogical inference is correct. Reasoners must consider uncertain theoretical reasons and empirical evidence for and against the validity of the inference.

Research As Developing Many Models of Many Studies

In real science, researchers do not settle an issue with a single study; they must develop models for many studies in which the same explanatory paths repeatedly appear. For example, to understand fermentation Buchner conducted many other experiments to rule out alternative hypotheses and to establish the postulated causal sequence. Some experiments used a method similar to the one shown in Figure 1; other experiments used quite different methods, such as warming the juice to observe what chemicals separated out. Other scientists using still different methods also contributed to the understanding of fermentation. Conducting many different experiments at different levels of analyses requires the construction of many models such as the ones in Figures 1 and 2.

The Need for More Authentic Tasks Based on More Complex Models

Our analysis has shown that authentic scientific research and simple inquiry tasks differ strikingly in required cognitive processes and underlying epistemology. These differences arise because authentic scientific research involves developing rich, complex models of data, whereas simple inquiry tasks are based on very simple models of data. Our analysis has suggested a need to develop new school tasks that come closer to reflecting the cognitive processes and epistemology of real science. These tasks should be based on models of data that are more complex than the models of simple experiments. Students should have experience constructing and evaluating research based on complex models of the type illustrated in Figures 1 and 2.

Part 2: Analyzing Current Inquiry Tasks

If, as we have argued, it is important to develop school inquiry tasks that do a better job of simulating authentic science, then a useful starting point is to examine a range of current inquiry tasks to see how close current tasks are to authentic science. In this section we apply our framework to analyze two groups of inquiry tasks: textbook inquiry tasks and inquiry tasks developed by researchers (including psychologists and educational researchers). We will address the following questions: To what extent do current textbook inquiry tasks incorporate features of authentic inquiry? To what extent do inquiry tasks developed by researchers incorporate features of authentic inquiry? Which features of authentic inquiry have been incorporated frequently into existing inquiry tasks, and which have been incorporated only rarely? What can we learn from existing tasks about how to create tasks that incorporate features of authentic scientific reasoning?

METHOD

To investigate the extent to which typical textbook inquiry tasks and inquiry tasks designed by researchers resemble authentic scientific inquiry, we analyzed 468 inquiry tasks in nine textbooks written for upper-elementary and middle schools and 26 inquiry tasks developed by researchers. We describe the analyzed samples below.

Textbook Tasks

We examined inquiry tasks in textbooks because textbooks continue to be influential in science classrooms (AAAS, in press; Stinner, 1995). We focused our textbook analysis on middle school and upper-elementary-school textbooks for two reasons. First, this is an age at which students are likely to develop strong beliefs about the nature of science (cf. Carey et al., 1989). To foster the development of an authentic scientific epistemology, it may be especially important for textbook inquiry tasks at this age to begin to incorporate key features of authentic scientific inquiry. Second, we intended this analysis to complement the analysis of Germann, Haskins, and Auls (1996), who investigated the extent to which laboratory tasks in high school textbooks support genuine inquiry.

Our analysis was conducted on the following textbooks: *Concepts and Challenges in Life Science* (Bernstein et al., 1991), *Heath Life Science* (Bierer & Lien, 1984), *Holt Physical Science* (Ramsey et al., 1986), *Life Science*: *Challenge of Discovery* (Warner et al., 1991), *Macmillan Earth Science* (Danielson & Denecke, 1986), *Merrill Life Science* (Daniel, Ortleb, & Biggs, 1995), *Merrill Physical Science* (Thompson, McLaughlin, & Smith, 1995), *Science Horizons* (Mallinson et al., 1993), and *SciencePlus*® (McFadden & Yager, 1993). These textbooks included 468 hands-on science activities (ranging from 28 to 84).

Researcher-Designed Tasks

Psychologists and educational researchers have developed a variety of tasks to investigate scientific reasoning. The earliest tasks were extremely simple tasks that bore very little resemblance to authentic science, such as the 2-4-6 task. In the 2-4-6 task, reasoners are told that there is a rule for generating sequences of three numbers and that the sequence 2-4-6 is an instance of the rule. Individuals try to figure out what the rule is by proposing additional sequences of three numbers (e.g., 1-3-5 or 2-12-18) and getting feedback on their "experiments." Clearly, this task bears little resemblance to authentic scientific research. Later tasks, however, have incorporated more and more features of authentic science. Table 3 briefly summarizes the 26 tasks that we analyzed. This list was intended to be a representative cross section of major tasks that have recently been developed by researchers. Almost all of the tasks have been described in published reports during the past decade. Several different kinds of tasks were included, including hands-on inquiry tasks (HOI), computer-simulated experimentation tasks (CSE), tasks involving the analysis of databases (DB), tasks evolving the evaluation of evidence presented as research reports (EE), and tasks in which people are asked to verbally design research studies (VDR). Most of the tasks used topics in the biological or physical sciences. Two important recent tasks have used psychological topics, and these were included, as well.

Coding

We coded each of the textbook tasks and each of the tasks developed by researchers according to the features displayed in Table 4. Eleven of the 14 features in Table 4 are features of authentic science. The remaining three features (simple control of variables, simple transformation of observations, and multiple studies of the same type) were included because these three features proved to sharply distinguish researcher-developed tasks from textbook tasks.

Table 4 provides a brief definition and an illustration of each feature included in the analysis. For textbook tasks, we judged whether each feature in Table 4 was reflected in the textbook's written description of the task. For the researcher-developed tasks, we based our judgments of whether each task incorporated the features in Table 4 on two sources of information: published descriptions of the tasks and research reports describing students' interactions with the tasks.

RESULTS

Table 4 presents the results of the analyses. Overall, the inquiry tasks developed by psychologists and educational researchers incorporated many more features of authentic reasoning than the textbook tasks. Despite differences in publishers and topics (earth science, physical science, and life science), the research activities in the nine textbooks

TABLE 3
List of Researcher-Developed Tasks That were Included in the Analysis

Name of Task	Type of Task	Description	References
Astronomy village	CSE	Using a computer simulation, students investigate astronomical questions such as whether an observed stellar event is a supernova.	NASA Classroom of the Future, 1996
BGuILE: Finch scenario	DB	Using a large database of information about the finches, animals, plants, habitats, and climate in one of the Galapagos islands, students try to figure out why some finches have shown a large decrease in population.	Reiser et al., 2001; Sandoval & Reiser, 1997; Tabak et al., 1996
BigTrak	HOI & CSE	Students work with a programmable toy vehicle to figure out the function of the "repeat" command.	Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993
Birds of Antarctica	DB	Students conduct investigations using a database containing information about 26 observed species of birds as well as meteorological and oceanographic information recorded during an actual 1982 voyage of scientists to Antarctica.	Maor & Taylor, 1995
Boat design	HOI	Students design small boats, trying to maximize their carrying capacity.	Schauble et al., 1995
Canal	HOI	Reasoners work out which of several variables influence the speed of a model canal boat as it travels down a trough filled with water.	Schauble, 1996; Schauble, Klopfer, & Raghavan, 1991; Zohar, 1995
Cancer drug trials	CSE	Medical students conduct simulated Phase I trials of a cancer drug.	Hmelo et al., 1998; Hmelo-Silver, Nagarajan, & Day, 2002, this issue
Car	CSE	Reasoners work out which of five variables influence the speed of a simulated car on a computer screen.	Kuhn et al., 1995; Schauble, 1990
Dinosaur extinctions	EE	Students evaluate theories about why the dinosaurs became extinct at the end of the Cretaceous period.	Cavalli-Sforza, Weiner, & Lesgold, 1994; Toth, Suthers, & Lesgold, 2002, this issue
Earthquake-resistant construction	HOI	Reasoners try to understand how to construct buildings made of blocks that will withstand the shaking of a	Azmitia & Crowley, 1998

TABLE 3
List of Researcher-Developed Tasks That were Included in the Analysis *(Continued)*

Name of Task	Type of Task	Description	References
		platform,which simulates an earthquake.	
Ecozone	HOI	Students investigate relationships between biotic and abiotic features of plots of land on the school grounds.	Roth & Bowen, 1993
Genetics construction kit	CSE	Provided with a population of organisms with specified traits and variations among those traits, students develop a theory that accounts for observed patterns of inheritance.	Hafner & Stewart, 1995; Stewart et al., 1992
GenScope	CSE	Students investigate genetic principles by designing dragons.	Horwitz, Neumann, & Schwartz, 1996
Geometric shapes	CSE	On a computer display, reasoners fire particles at geometric shapes objects as they try to discover the rule that governs the motion of the particle.	Klayman, 1988; Mynatt, Doherty, Tweney, 1977
Hunger in the Sahel	CSE	By determining how to employ 10 lots of land that differ in steepness, students try to maintain an agricultural profit when also faced with varying climate conditions.	Leutner, 1993
Lakes	CSE	Reasoners investigate how the population of four species of sea animals in a tank are influenced by four variables: temperature, salt concentration, oxygen concentration, and the current.	Vollmeyer, Burns, & Holyoak, 1996
Microbiology experiment design	VDR	Students describe verbally how they would set up an experiment to test a hypothesis in microbiology.	Dunbar, 1999
Objects falling through water	HOI	Reasoners investigate the factors that influence how fast objects fall through a tube of water.	Penner & Klahr, 1996
Pond ecology	HOI	Students investigate the effects of pollution in a local pond.	Rosebery, Warren, & Conant, 1992
Rocket flight	HOI	Students work out variables that influence how high a rocket flies.	Petrosino, 1999

Continued

TABLE 3
List of Researcher-Developed Tasks That were Included in the Analysis *(Continued)*

Name of Task	Type of Task	Description	References
Simulated molecular genetics lab	CSE	Reasoners figure out the mechanism by which genes regulate the production of the enzyme that converts lactose to glucose.	Dunbar, 1993; Okada & Simon, 1997; Schunn & Dunbar, 1996.
Simulated psychology lab	CSE	Reasoners test two theories of the “spacing effect” in human memory. Reasoners test these theories by conducting simulated verbal learning studies in which simulated subjects study lists of words under various conditions.	Schunn & Anderson, 1999
Tank system	CSE	Reasoners try to control the temperature of four interconnected water tanks.	Moray, Lootsteen, & Pajak, 1986
Taste of cola	VDR	Subjects describe verbally how they would set up and conduct a gustatory experiment to find out what people like about Coca Cola’s taste.	Schraagen, 1993
ThinkerTools	CSE	Students conduct experiments involving motion of objects.	White & Horwitz, 1988; White & Frederiksen, 1998
WISE: Deformed frogs mystery	EE	Using resources found on the project’s internet pages, students read about and evaluate evidence relating to issues such as the cause of observed deformities in many frogs.	Linn, Bell, & Hsi, 1998

Note: HOI = hands-on inquiry tasks; CSE = computer-simulated experimentation; DB = database investigations; EE = evidence evaluation tasks, VDR = verbal designs of research.

consistently failed to incorporate elements of authentic scientific reasoning (an average of less than 0.5 of the 11 features of authentic science per task). The 26 tasks developed by researchers included more authentic features (2.9 of 11 features per task). These tasks included 15 tasks that have been used in classrooms (3.4 features per task) and 11 tasks that, to our knowledge, have been used only in laboratory studies (2.3 features per task).

We found no textbook activities in which students were allowed to generate their own research question. This feature was also uncommon in the researcher-generated tasks. In only 2% of the textbook activities were students allowed to select their own variables to investigate. This feature was common in researcher-developed tasks (50%).

In textbook tasks, there were few opportunities for students to think about controlling variables. Far from worrying about issues of complex control in authentic experiments, students were seldom asked even to consider how to control variables in simple ways (4%

TABLE 4
Features of Authentic Inquiry in Textbook and Researcher-Developed Tasks

Feature of Reasoning Task	Definition and Example	Textbook Tasks	Researcher-Developed Tasks
Generating own research question	Learners are not told what question(s) to investigate but develop these questions on their own, e.g., learners are told to investigate a question of their own about life in a 35 m^2 plot of land.	0%	12%
Selecting own variables	Learners are not told exactly what the relevant variables are but select and/or define these variables on their own, e.g., learners are directed to think of and investigate variables that might influence the flight of paper airplanes but are not told what these variables are.	2%	50%
Developing simple controls	Learners must control already-known variables, e.g., learners investigate the effects of engine size, wheel size, color, and two other variables on car speed; when investigating the effects of one variable, the other four variables must be held constant. Note: Many textbook tasks control variables implicitly when they direct learners to carry out two different procedures and compare the results. However, learners themselves do not ever consider issues of control; they merely follow directions.	4%	92%
Developing relatively complex controls	Learners must be concerned about nonobvious controls, e.g., learners must devise a way to control for the amount of light shining on two plots of land that are differentially shady, or learners propose a counterbalanced design when designing a psychology experiment.	1%	27%
Making multiple observations	Learners measure or evaluate measures of multiple variables, such as observing several different parts of a plant or observing population changes in several different species in a lake ecology simulation.	17%	54%
Observing intervening variables	Learners measure or evaluate measures of intervening variables,	4%	38%

Continued

TABLE 4
Features of Authentic Inquiry in Textbook and Researcher-Developed Tasks *(Continued)*

Feature of Reasoning Task	Definition and Example	Textbook Tasks	Researcher-Developed Tasks
	e.g., learners examine the ways in which bird behavior as an intervening variable mediates the effects of drought on bird survival.		
Using analog models	Learners conduct research with simplified analog models intended to represent real situations, e.g., learners experiment with rocks and sand in a jar to model sediments in the ocean.	15%	15%
Simple transformation of observations	Learners transform observations in simple ways such as averaging data and/or graphing results.	2%	42%
Complex transformation of observations	Learners transform variables in ways that go beyond averaging or graphing, e.g., learners analyze telescope images of several regions of space, and then use an image processor to make movies of the images in order to determine whether any spots of light change in brightness.	0%	12%
Consideration of methodological flaws	Learners reason about possible experimental flaws in the method of the study they are designing or interpreting, e.g., learners worry about whether a method for measuring sunlight in a 1-m^2 plot of land is accurate, or learners note possible flaws in the methods used by scientists to gather data about penguins in Antarctica.	2%	19%
Developing theories about mechanisms	Learners develop or test theories about mechanisms, e.g., learners develop theories about how genes regulate other genes, or learners test two rival theories of why spaced review is effective at enhancing memory.	0%	35%
Multiple studies of the same type	Learners conduct more than one study as they engage in inquiry on a topic, and the studies are all of the same type, e.g., learners conduct many studies on factors that influence how fast toy boats travel, but all the	2%	81%

TABLE 4
Features of Authentic Inquiry in Textbook and Researcher-Developed Tasks *(Continued)*

Feature of Reasoning Task	Definition and Example	Textbook Tasks	Researcher-Developed Tasks
	studies involve the same basic procedure of running a boat down a canal.		
Multiple studies of different types	Learners conduct different types of studies, such as clinical drug trial studies and studies using cell cultures.	1%	23%
Studying expert research reports	Learners read research reports written by scientists or abbreviated, newspaper- or magazine-style reports of such research.	0%	12%

of activities). Simple control of variables was very common in researcher-developed tasks, and complex control of variables appeared in about a quarter of these tasks. For example, in several tasks students must tease apart the effects of intercorrelated independent variables. Still, we think that more can be done in the area of complex control. Common techniques of more complex control such as counterbalancing, conducting blind tests, and running external controls have not yet appeared in reasoning tasks that could be used instructionally in schools.

About 17% of the textbook tasks incorporated multiple observations. However, almost all instances of multiple observations occurred in simple observations, when students were directed to observe several aspects of an object such as a starfish or a plant. There was almost no use of multiple observations in simple textbook experiments or simple textbook illustrations. Thus, the use of multiple observations was limited to a very restricted subset of activities. The use of multiple observations was much more common in the researcher-developed tasks. Moreover, more than a third of the researcher-developed tasks involved observations of variables such as intervening variables.

Textbook activities incorporated analog models with some regularity (15% of all activities). For example, in one activity students place gravel, small pebbles, sand, soil, and water in a jar to investigate the process of sedimentation. The processes inside a jar are assumed to be analogous to the processes that occur in a lake or ocean. However, it should be noted that although the textbooks employed analog models, they did *not* encourage students to reflect on the soundness of the models. For instance, in the sedimentation task, students were simply told that the processes inside the jar were analogous to the processes in the ocean, but they were not asked to consider the validity of this analogy. Use of analog models occurred in 15% of the researcher-developed tasks; in contrast to the textbook tasks, these tasks did encourage students to consider whether the research model was a sound analog of the real situation.

In the textbook tasks, there was little transformation of data, no explicit concern with possible bias in observations, and little concern with experimental flaws. Simple transformations of data occurred more frequently in the researcher-developed tasks (42%), but more complex transformations were infrequent (12%). In addition, few researcher-developed tasks encouraged students to be concerned with methodological flaws. In most of these tasks, learners were apparently expected to assume that the data are reliable. The constant worry about possible methodological error that pervades scientists' work is not present in most of these tasks. This conclusion is supported by the published reports of students engaged in these tasks, which only rarely describe learners expressing concerns about methodologies.

One feature of authentic science that has been incorporated into more than one-third of the tasks developed by researchers is that learners are asked to develop theories involving mechanisms. Textbook tasks generally direct students to investigate surface-level features of the world such as size, weight, distance, speed, and color. If theories are involved at all, students are told exactly what the theories are and how they are linked to the world. By contrast, in nine of the tasks developed by researchers (35%), students investigated underlying theoretical mechanisms such as how the parental genes combine to affect the observed traits of offspring. Students engaged in these tasks must use indirect reasoning to connect theoretical variables to observed variables.

Textbook tasks only rarely asked students to conduct multiple studies of any kind. By contrast, researcher-developed tasks regularly expected learners to conduct multiple studies; however, these studies were usually of the same type rather than of different types.

Finally, students using textbook tasks did not ever read real research reports. Students read research reports in only three of the researcher-developed tasks (12%).

DISCUSSION

One conclusion to be drawn from this study is that the inquiry activities in most textbooks capture few if any of the cognitive processes of authentic science. As a result, textbook inquiry tasks assume an epistemology that is entirely at odds with the epistemology of real science. There is no coordination of theory with complex sets of partially conflicting data. The theory-ladenness of methods is not at issue, nor are students encouraged to think about alternative interpretations of the data they generate. The reasoning is algorithmic, as students draw obvious inquiry conclusions from simple experiments and simple observations.

The results of our analysis of 468 activities in nine middle-school and upper elementary textbooks are similar to the results of a recent analysis of 90 laboratory activities in nine high school biology texts by Germann, Haskins, and Auls (1996). Our results are not directly comparable to the results of Germann, Haskins, and Auls, because they employed a different theoretical framework. Most of their categories were categories on which even simple inquiry tasks could potentially score highly. For instance, Germann, Haskins, and Auls coded for whether tasks required students to formulate hypotheses, design observations, design experiments, control variables, and provide evidence. These are processes that could occur in simple as well as complex inquiry tasks. However, the results of Germann, Haskins, and Auls indicate that most inquiry activities did not even include these simple reasoning processes. For instance, only 13.3% of the activities required students to formulate hypotheses, only 4.4% expected students to design even one experiment, and only 4.4% required students to control variables. Only 1 of 90 activities asked students to identify independent variables. Thus, despite the use of different theoretical frameworks, the overall picture obtained by Germann, Haskins, and Auls nonetheless agrees with ours: Textbooks seldom engage students in any kind of real inquiry, not even good simple inquiries, let alone more authentic inquiries. Together, the two analyses suggest that far from promoting an authentic epistemology of science, textbooks may in fact promote an inauthentic view of science as a process of accumulating simple facts about the world.

Our analysis suggests that several recent tasks developed by researchers have done well at capturing one central epistemological feature of science—building and revising theoretical models. Many researcher-developed tasks have also moved a great distance away from algorithmic reasoning, and several tasks engage students in highly uncertain reasoning. It is less common in researcher-developed tasks to ask students to coordinate results from different types of studies. The theory-ladenness of data is absent from most tasks, as is encouragement of a full range of responses to anomalous data that includes rejecting on methodological grounds. The institutional aspects of the social construction of knowledge are also absent from most tasks.

We think that the differences between textbook tasks and researcher-developed tasks arise from differences in the models of data that underlie these tasks. Textbook tasks are based on very simple models such as the ones in Figure 3. Researchers have improved upon textbook tasks by developing tasks with more complex underlying models. More complex models afford greater use of authentic scientific reasoning processes and thus encourage a more scientific epistemology.

Our analysis leaves open the question of how similar textbook inquiry tasks are to other inquiry tasks in science education. We think that the simple inquiry tasks that we found in science textbooks are representative of the inquiry tasks that are found in many other science education materials, including trade books of science experiments (e.g., Murphy, 1991; Penrose, 1990; VanCleave, 1997; Walpole, 1987; Whalley, 1992) and commercially produced educational software (e.g., Houghton Mifflin Interactive, 1997; Theatrix Interactive, 1995). However, our analysis leaves open the question of how authentic the inquiry activities are in more innovative curricula, such as the highly regarded Nuffield courses. We would expect that there are many innovative curricula with inquiry tasks that come much closer to authentic inquiry than textbook inquiry tasks do. Still, our analysis of recent tasks developed by researchers shows that there is still room for improvement even in these outstanding, cutting-edge inquiry tasks, we would conjecture that there is room for improvement in the best existing inquiry curricula, as well. The important point is that our framework can serve as an analytic tool to help educators evaluate inquiry tasks in such innovative curricula to find out which features of authentic inquiry are incorporated and which are not.

The usefulness of the framework in evaluating innovative inquiry tasks is illustrated by examining an exemplary inquiry activity for fifth graders described in the National Research Council’s recent addendum to the U.S. National Science Education Standards (Olson & Loucks-Horsley, 2000). In this activity, fifth graders decided to investigate a puzzle raised by a row of three trees on the school grounds. The tree at one end of the row had healthy leaves, the tree at the other end had yellowing leaves, and the tree in the middle had a mixture of healthy and yellowing leaves. The students investigated why only one of the trees was thriving, investigating variables such the age of the trees, the amount of water, water contamination, and insect infestation. The students studied books on trees and their life cycles and conducted systematic observations. After making many observations for several weeks, often at hourly intervals, the students concluded that the unhealthy trees were getting too much water. We agree that this is an exemplary task for fifth graders that captures more of the complexity of authentic science than almost all of the textbook tasks that we examined. For instance, students generated their own research question, selected their own variables to investigate, and read about trees in books (although they did not read about real research). Students coordinated a fairly complex array of evidence to notice patterns of correlation. However, using Tables 1 and 2 as a guide, we notice that there are other features of authentic science that appear to be are absent, such as consideration of methodological flaws, measurement of intervening variables, design of studies at different levels of analysis, and the development of underlying theories. Of course, there is no need for every inquiry task to incorporate every feature of authentic science. It is also possible that some features of authentic science are better left for middle school or high school; this is an empirical issue that requires further research. Our main point is that the use of the framework provides a method for analyzing the inquiry tasks in a curriculum to determine if they collectively incorporate all aspects of scientific reasoning.

FIVE TYPES OF REASONING TASKS

In this section we take a closer look at several of the very promising reasoning tasks in Table 3 in order to gain insights into how to develop more authentic reasoning tasks for schools. The inquiry tasks presented in Table 3 can be classified into five basic categories: hands-on inquiry, computer-simulated experimentation, database tasks, evidence evaluation tasks, and verbal designs of research. Each of these task types has inherent strengths and limitations.

Hands-On Inquiry

In hands-on inquiry, students conduct investigations with real-world materials. Hands-on inquiry tasks can range from capturing no features of authentic science to capturing many features of authentic science.

The simple experiments found in textbooks are a prominent form of hands-on inquiry, and these tasks, as we have discussed at length, share few if any of the features of authentic scientific research. The tasks assume very simple models such as the one in Figure 3, which are so oversimplified that there is little real science left.

Hands-on inquiry comes much closer to authentic science in relatively free inquiry tasks. The ecozone task described by Roth and Bowen (1993) is a good example. In this task students worked in pairs and were assigned ecozones of 35 m^2 to investigate. Students' task was to find out about and report the relationships between biotic and abiotic features of their ecozone. Students could choose their own questions, and they were given access to equipment such as soil corers, soil moisture meters, and pH meters. An example of a student project was one dyad's investigation of the relationship between soil pH and the density of plants. The two students decided to stake out three 1-m square plots and then measure soil pH and the density of plants in each square. Later, they became interested in measuring light intensity in their plots, as well. However, the students noticed that the light across each of the squares was uneven because of shadows partially covering each region. As a result, they began to worry about how to measure light intensity. One student thought that it would be sufficient to make one measurement in the middle of each square, whereas the other thought that they needed to measure the light in the middle of each square and at each of the four corners to gain a more complete picture of the light in each square. In addition, the students discovered that it was problematic to decide how high to place the light meter in these five spots. Should the meter be at ground level or at a higher level, more on the level of the plants' leaves? In this case study, students exhibited concern with the reliability of their methods, and they struggled over how to achieve control over relevant variables.

In this free inquiry task, we see hands-on inquiry at its best. Free inquiry tasks have the potential to incorporate several key features of authentic scientific reasoning. Students are free to construct more complex models of experiments as they conceptualize their studies. Students can worry about appropriate methods, about whether measures are biased, and about how to control for complex confounds. However, it is unlikely that students will exhibit such concerns without encouragement from teachers.

It appears to us that one inherent limitation of free hands-on inquiry in the classroom is that it is relatively difficult for students to conduct experiments at deep theoretical levels of analysis. For example, although students can investigate relationships between soil pH and plant growth, it would be difficult in the classroom to conduct biomolecular investigations of

TABLE 5
Strengths and Limitations of Five Types of Simulated Research Tasks

Task Type	Strengths	Limitations
Hands-on inquiry	Free hands-on inquiry is especially good at encouraging invention of students' own variables, complex control of variables, and consideration of methodological flaws.	Even with free hands-on inquiry, it is difficult to conduct different kinds of experiments, especially experiments at a theoretical level. Studies tend to address the level of observable phenomena.
Computer-simulated experiments	Computers permit students to conduct experiments of different types quickly, including experiments focused at the theoretical level. Designs can be complex, and systems can potentially simulate methodological flaws.	It is difficult to simulate complex control of variables, and learners cannot invent theories or variables that the system has not anticipated. Learners are asked to make choices about variables they might not consider on their own.
Databases	Database tasks have the potential to simulate most features of authentic reasoning in the context of interpreting existing data.	Database tasks do not allow students to design and carry out their own studies and do not permit new investigations on some unanticipated issues that arise during inquiry.
Evidence evaluation	Evidence-evaluation tasks have the potential to simulate most features of authentic reasoning in the context of comprehending and interpreting existing studies.	Evidence evaluation tasks do not allow students to design and carry out their own studies and may not permit new investigations on unanticipated issues that arise during inquiry.
Verbal design of studies	Verbal designs of studies simulate most features of authentic reasoning in the context of describing a design for a study.	This task is more suitable for assessment and research than instruction, because students have no opportunity to actually implement designs and get feedback from the environment. There is also no assessment of interpretive skills.

why acidic soil affects different plants in different ways. Hands-on inquiry in the classroom will tend to focus on investigating observable phenomena rather than on testing underlying theories.

Computer-Simulated Experimentation

Many of the tasks listed in Table 3 are computer-based simulations of experimentation. A good example is Dunbar's molecular genetics lab (Dunbar, 1993). In this simulation, students' goal is to understand how genes regulate the production of the enzyme β-gal, which converts lactose into glucose. Students are shown a display with a sequence of six genes, represented by a series of six adjacent squares. The first three genes are labeled I, P, and O. The last three are unlabeled, and students are told that these three unlabeled genes are responsible for producing β-gal. Students set up experiments by varying (a) whether to run an experiment with haploid cells (just one set of the I, P, and O genes) or with diploid cells (having two sets of each of these three genes); (b) what kind of mutations of the I, P, and O genes to use; and (c) how much lactose to add to the cells. When students run experiments, the outcome variable is how much lactose is present in the cells. Through conducting experiments, students can work out that the I and O genes exert a suppression effect that inhibits production of β-gal by the three unlabeled genes. In this task, students must employ indirect paths of reasoning to develop theoretical mechanisms that can explain observations.

When compared with hands-on inquiry, computer simulations offer an important advantage but suffer from a serious drawback. The advantage is that computers allow students to conduct simulated experiments with complex underlying models that they could not conduct in reality because of lack of time and equipment. This allows computer-simulated experiments to capture several features of authentic reasoning that are hard to capture using hands-on inquiry. First, computers allow students to conduct experiments at the level of theoretical mechanism. For instance, the molecular genetics lab permits students to investigate the details of gene control, which would be very difficult to do in the classroom. By partially reducing the complexity of real experiments and by simulating the use of expensive equipment, computer simulations permit students to investigate theoretical entities.

A second feature of authentic science that can be captured easily by computer simulations is the use of different types of experiments. One can easily envision a computer simulation environment in which students can conduct different types of experiments on the same issue. For example, the causes of a disease could be probed in one kind of simulated study using cell cultures, a second type of simulated study using mice, and a third type of simulated study involving clinical trials of a medicine.

A third feature of authentic reasoning that can be incorporated into computer simulations is the possibility of implementing relatively complex designs. The cancer drug trial simulation described by Hmelo-Silver et al. (2002, this issue) employs relatively complex, multistep procedures that capture a great deal of the complexity of real experimentation.

Computer simulations could also be designed to simulate experiments in which methodology is a major concern, although they have not yet to our knowledge been designed in this way. One can imagine a simulation in which the learners use different methods to investigate an issue, and these methods yield conflicting results, which would impel learners to think about how to reconcile the rival methods or how to decide which is more reliable. This is a promising direction for the future development of computer-simulated experimentation.

Against these virtues of computer-simulated experimentation, there are two limitations. First, in computer simulations the variables must be largely predefined, and a great deal of the messiness of the natural world is artificially cleaned up. In order to run the simulations, the computer must be programmed in advance to "know" what the causal variables are, along with all potential alternative models. There is no way to allow students to test alternative models or novel variables that were not programmed into the system.

A related limitation of computer-simulated experimentation is that computers require students to make choices about every variable that the computer considers relevant. In computer-simulated experiments, students choose from a menu of variables, and they must make a decision about every variable on the menu when designing every experiment. To give an example, when students conduct a hands-on experiment on the effects of light on seed germination, they may simply not think about certain relevant variables that need to be controlled, such as the depth of the seed, the distance from the corners of the container, and even the type of seed. A computer simulation generally requires students to select the value of each variable for each trial (e.g., a student is required to choose the type of seed and depth of planting), which makes the control of these variables salient. The computer requires the learners to attend to variables that they might not have noticed on their own. As a result, student may not learn to control variables in situations where they are not presented with a priori lists of variables.

Databases

Databases differ from computer-simulated experimentation in that students using computer-simulated experimentation design experiments and "gather data" during the simulation, whereas students using databases examine evidence that has already been gathered. An example of a database comes from Reiser's BGuILE project (Reiser et al., 2001; Tabak et al., 1996). BGuILE is a set of computer-based investigations designed to facilitate students' learning about several topics in biology, including evolutionary change. In one of these tasks, students' goal is to explain why some (but not all) finches on one of the Galapagos islands were dying during 1977. Students use the computer to examine data of their choice. Students can examine measurements of individual birds, observe finch behavior on video clips, investigate weather data, and examine the features of the habitats. One tenable solution to the problem involves noticing that there was a drought in 1977 and that the drought nearly destroyed the population of plants with small seeds. One plant that was hit less hard had large but thorny seeds. The finches with longer beaks were better able to penetrate the large, thorny seeds of this plant, and so it was the finches with longer beaks that were more likely to survive. It is apparent that as students investigate this question, they must construct complex causal models that integrate elements of finch anatomy, finch behavior, weather, habitats, the presence of other animals, and so on.

The finch task captures many features of authentic scientific reasoning. Students grapple with real data about birds, climate, plants, and behavior and try to make sense of it. Students have some scope to invent complex variables, although it is true that many of the variables to be investigated are provided in advance (such as rainfall in a given year). The procedures themselves are not complex (checking weights of birds or finding out amount of rain is straightforward), but the need to integrate many different sources of information is great. Conceptual control of variables is important. For example, students must be sure that the correlation between beak size and survival during the drought is not confounded with correlations with any other variables. Students must transform raw data (e.g., individual birds' beak measurements) into usable generalizations (e.g., birds with large beaks eat thorny seeds) that can be explained using a theoretical explanation (e.g., explanations in terms of selective pressures). Students must use indirect reasoning strategies to link explanations employing evolutionary concepts to observed phenomena. As in real science, the data are potentially amenable to different interpretations, so that no one conclusion is ever certain.

Databases such as the BGuILE finch scenario do not permit the full scope of scientific inquiry, because students can only decide what data to examine and are not free to gather any data they wish. However, databases can provide a rich, complex set of information that requires highly complex reasoning. Databases can include many irrelevant as well as relevant variables. Many of the relevant variables may be intercorrelated, which requires students to develop strategies for teasing apart the effects of associated variables. To the extent that databases provide detailed information about how the data in the various parts of the database were gathered, students can begin to question some of the methods used. Databases can also potentially provide information at different levels of analysis. For instance, a database containing data relevant to whether bacteria cause ulcers could include data from epidemiological studies, data from clinical trials of administering antibiotics to ulcer patients, and microscopic photographs of stomach samples of various patients. Thus, although databases do not involve students in designing research studies, they do capture many aspects of interpreting data and developing theories. They also reflect an authentic mode of doing science, as internet-based databases are becoming important in many fields of science (Thagard, 1999).

Evidence Evaluation Tasks

Evidence evaluation tasks present students with written reports of evidence and ask them to draw conclusions from them, as scientists do as they read and discuss evidence bearing on rival theories. The WISE project (Bell & Linn, 2000; Davis & Linn, 2000; Linn, 2000; WISE, n.d.) is one example of an evidence evaluation task. WISE consists of a series of internet-based investigations such as trying to understand why frog populations in North America have been deformed, determining how best to fight malaria, and predicting the next earthquake in California. To take one example of an investigation that has appeared on the WISE website, the deformed frogs mystery presents students with evidence showing that many frogs have recently been born with additional or deformed legs. Students consider whether parasites or environmental chemicals are the causes of the deformities. They find a variety of evidence on the web site. The web site's own pages present brief summaries of studies, and the site contains links both to newspaper articles that report additional research and to scientific articles posted on sites such as the web site of the U.S. National Institutes for Health. The studies include controlled lab experiments as well as demographic evidence such as maps that show the incidence of frog deformities in counties throughout the United States.

Evidence evaluation tasks such as WISE differ from previously discussed tasks in that students are not just presented with data; they are also presented with scientists' explanations for the data. For instance, one study in the deformed frogs mystery presents a picture contrasting a naturally deformed frog leg with a leg deformed by exposure to a class of chemicals called retinoids. In the verbal description of the results, the scientist explicitly concludes that the best explanation for the frog deformities is exposure to retinoids. This feature of WISE sets WISE apart from most other tasks, which present raw data and refrain from presenting scientists' interpretations. WISE thus fills an important niche among tasks that attempt to simulate authentic science. It places students in the position of lay adults who must make policy decisions based on conflicting scientific reports that include theoretical interpretations along with data.

Like databases, evidence evaluation tasks hold promise for capturing a range of features of authentic reasoning. Although pure evidence evaluation tasks do not place students in the position of needing to design studies, students do have to evaluate the interpretations they read, judge whether there are alternative interpretations that could be substituted for the provided interpretations, and coordinate conflicting evidence. They must comprehend and evaluate studies that may have quite complex underlying models. The studies can involve many intervening variables with multiple measures. Students must coordinate results from many different kinds of studies in order to develop a tenable theory. As with databases, if sufficient information is given about the methods used in various studies, students can reflect on the validity of the various methodologies. Evidence evaluation tasks, at their best, place students in the position of scientists who are reading a wide range of evidence and making up their minds about what to believe.

Verbal Design of Research

In all the tasks we have discussed so far, reasoners gather and/or examine data. Schraagen (1993) used a very different approach to examine reasoning. Schraagen presented Dutch adults with a scenario in which the Coca Cola company was interested in sponsoring a study that would be used to aid in marketing Coca Cola in the Netherlands. The participants' task was to explain how they would design a study to determine exactly what it is that people taste when they drink Coca Cola versus Pepsi Cola and a house brand.

In comparison to the previous tasks, an obvious drawback of this task is that there is no opportunity for participants to examine or interpret data. However, a virtue of the task is that as people design their experiments, they are faced with all the complexity of designing a real experiment. They must construct a complex model for an experiment just as scientists do. They must develop complex procedures and consider what kind of experimental task would be appropriate to predict consumption of cola by the Dutch population. They must consider complex issues of control such as how to counterbalance the order of tasting different colas, how to eliminate lingering taste when tasting one cola after another, and how to preclude experimenter bias. They can introduce multiple measures such as rating scales, forced choice measures, and so on. They may consider how to generate a representative sample, and they must worry about possible methodological flaws and potential sources of bias in the observations.

Dunbar (1999) has recently used a verbal design task in which undergraduates are asked to design experiments in the domain of biology. His task requires that reasoners propose indirect methods to get at underlying mechanisms. Thus, verbal design tasks can also be used to find out what people know about designing studies to test theoretical mechanisms.

In short, verbal designs of experiments require reasoners to consider many different aspects of authentic experimentation as they describe how they would design an experiment. In contrast to computer or other simulations, which tightly constrain the parameters of the inquiry task, verbal descriptions of how to design an experiment can address any of the real-world complexities of research. Unlike the other types of simulation tasks, however, this task type does not permit the researcher to investigate how reasoners coordinate results from multiple experiments or how they interpret actual results from any experiment. Verbal design tasks are limited as a classroom activity, but they hold promise as a means for psychologists and educational researchers to assess students' ability to reason.

Hybrid Systems

It is obvious that a good way to develop school reasoning tasks that capture all or most features of authentic science would be to combine two or more of the task types described in this section. Several recent research efforts have joined two or more forms of inquiry in hybrid combinations. To give just two examples, the WISE (e.g., Bell & Linn, 2000) and BGuILE research teams (Loh et al., 2001) have developed investigations that combine multiple forms of inquiry such as evidence evaluation and hands-on-inquiry. When two or more forms of inquiry with complementary strengths and weaknesses are combined, most features of authentic inquiry can be covered.

IMPLICATIONS

We think that our analysis of scientific reasoning has important implications for research, assessment, and instruction.

Research

Tasks developed by researchers to study scientific reasoning have made large strides over the past several decades. The earliest tasks incorporated few features of authentic reasoning, but later tasks have begun to incorporate progressively more features. We think that this has been productive as a strategy for psychological research. The earlier studies provided a good foundation for understanding how students reason about simpler tasks. With this foundation, researchers can expand on earlier studies to investigate more and more complex forms of reasoning.

We would suggest that the framework presented in this paper can be used to identify additional features to build into inquiry tasks for future research. Tables 1 and 2 can be used to identify dimensions along which new features of authentic reasoning can be added to inquiry tasks. For example, there is a particular need to understand how people reason when methods are in question and how people coordinate results from different kinds of experiments. There is a similar need to understand how students interpret and evaluate research reports.

Another important area for future research is to investigate inquiry tasks in nontextbook science curricula. Our analysis has shown that the typical textbook inquiry task captures none of the features of authentic science that we have discussed. What about inquiry tasks in science curricula that are not based on formal published textbooks. We suspect that analyses of inquiry tasks in these curricula will show that further improvement is possible, just as improvement is possible in tasks developed by researchers. But there will undoubtedly be exemplary inquiry tasks in future curricula that can serve as models for designers of the scientific reasoning of inquiry tasks.

Most research on how children and adults perform on scientific reasoning tasks has employed hands-on inquiry tasks and computer-simulated experimentation. Our analysis also suggests that there is a need for more research that employs database tasks, evidence evaluation tasks, and verbal designs of studies.

Another implication of the framework presented in this paper is that there is a need to gain a better understanding of the actual reasoning strategies used by scientists as they engage in authentic inquiry. The frameworks presented in Tables 1 and 2 point to many differences between scientific reasoning and the reasoning needed for most school tasks, but they fall short of providing a detailed account of the actual strategies that scientists use. For example, our analysis points out that scientists reason under uncertainty and that they coordinate partially conflicting results from different experiments of various types. However, our analysis stops short of saying exactly what strategies scientists use when they reason under uncertainty and when they coordinate conflicting results. Detailed studies of scientists' actual reasoning are needed to understand the specific strategies that scientists use. Knowledge of authentic reasoning strategies that scientists actually use can inform decisions about what reasoning strategies students should learn.

Assessment

The analyses presented in this paper have important implications for assessment. Many assessments of what children learn about science have made simplistic assumptions about the nature of scientific reasoning. Many assessments employ tasks based on relatively simple models of data. These assessments examine relatively simple strategies such as controlling variables in simple situations. We would argue that assessments should also assess students' ability to reason about tasks that incorporate more of the features of authentic scientific reasoning listed in Tables 1 and 2. These tasks should be based on more complex models of data such as the ones in Figures 1 and 2. Assessments of scientific reasoning should provide information about issues such as how students reason about possible methodological error, how they coordinate results from different experiments, how they understand more complex issues of control, and how they understand reports of scientific research. Some recent work has made important strides in incorporating features of authentic reasoning into assessments (see, e.g., Gitomer & Duschl, 1995; Glaser & Baxter, 1997).

Instruction

The analysis presented in this paper indicates that many current school inquiry tasks bear little resemblance to authentic scientific reasoning. The most serious problem is that school tasks may actually reinforce an unscientific epistemology. Many current school tasks may encourage the belief that science is a simple, algorithmic form of reasoning; as a result, students are likely to fail to learn the heuristics scientists use to reason under uncertainty.

Our analysis suggests that students should have opportunities to work with more authentic tasks that have more complex underlying models. Many of the researcher-developed tasks described in this paper have made great progress toward accomplishing this goal. In addition, we think that a variety of new tasks should be developed that incorporate aspects of authentic reasoning that have been less frequently incorporated into existing tasks. For example, there has been little development of inquiry tasks that enable students to learn how to reason about methodological flaws or how to coordinate theories with multiple studies that may conflict with each other.

More complex inquiry tasks will take a substantial amount of classroom time, far more than the single class periods often devoted to science labs. Work with systems such as BeGUILE requires extended work over many days. There is no way to condense authentic scientific reasoning into a single 40- to 50-min science lesson. Learning authentic scientific reasoning will require a commitment by teachers and schools to spend the time needed to learn reasoning strategies that go beyond simple observation and simple control of variables.

Even if ample time is allotted to authentic reasoning tasks, there will be serious instructional challenges. At present little is known about how to foster complex reasoning. It is difficult enough for students to learn to control variables in simple situations. How are students to learn a large number of more complex strategies needed for more authentic reasoning? There is a pressing need for research that develops and tests instructional approaches for fostering the development of such complex strategies.

We suspect that many teachers will welcome authentic inquiry tasks. Many science teachers are probably aware that neat-and-tidy classroom experiments do not resemble the much messier research that they read about in newspapers, magazines, and journals. If the curriculum were changed to allow more time for inquiry, many teachers would probably be eager to incorporate authentic inquiry tasks in their classrooms. However, it will surely not be sufficient simply to develop authentic tasks and make time for students to use them. Teachers must also know about reasoning strategies that are effective with such tasks as well as effective instructional strategies to help students master these reasoning strategies.

CONCLUSION

An important goal of science education is to foster the development of epistemologically authentic scientific reasoning. The ability to reason well about complex models of data is essential not only for scientists but for nonscientists as well. All citizens need to be able to reason well about complex evidence such as evidence relating to health and medical decisions, evidence relating to social policies upon which citizens vote, or evidence relating to the best way to promote employee motivation and satisfaction. Learning an oversimplified version of scientific reasoning will not help on such real-world reasoning tasks. Indeed, when students learn an oversimplified, algorithmic form of scientific reasoning in school, they are likely to reject scientific reasoning as irrelevant to any real-world decision making.

The results of our study indicates that much work remains to be done to transform schools into places that nurture epistemologically authentic scientific inquiry. Textbook curricula, which remain important in many schools, are dominated by oversimplified inquiry tasks that bear little resemblance to authentic scientific reasoning. Most of the inquiry tasks developed by researchers have incorporated several additional features of authentic reasoning. The best tasks developed by researchers have incorporated a majority of features, but most still omit several key features of authentic science—such as a central focus on the theory-ladenness of data. Thus, work needs to be done to develop curricula with authentic inquiry tasks, not just hands-on inquiry tasks but also evidence evaluation tasks, database tasks, and computer-simulated experimentation. All of these tasks should be designed so that they meet the criteria for authentic experimentation laid out in Tables 1 and 2. This means, in essence, that the tasks must be based on complex models of data (as in Figures 1 and 2) rather than on simple models of data (as in Figure 3).

A trio of goals must be met to promote authentic scientific reasoning in schools. The first is to develop reasoning tasks that afford authentic reasoning. The second is to develop a better understanding of the strategies that scientists use when reasoning on such tasks. The third is to develop instructional strategies that ensure that students learn these authentic reasoning strategies when they engage in authentic inquiry tasks. The remaining three papers in this special section take important steps toward achieving these goals. All three papers present examples of tasks that incorporate significant features of authentic inquiry. The paper by Hmelo-Silver, Nagarajan, and Day investigates reasoning strategies that are effective when students work with a highly authentic computer simulation of cancer drug trials. Shimoda, White, and Frederiksen employ an interesting hands-on inquiry task in the domain of psychology, and they develop effective scaffolding for helping students learn the complex reasoning needed to succeed at complex inquiry. In the last paper in this special section, Toth, Suthers, and Lesgold develop innovative techniques for representing theories and evidence that help students learn to reason about a realistic evidence evaluation task. Collectively, these three papers develop tasks provide exemplars of inquiry tasks that begin to incorporate key features of authentic inquiry, examine strategies needed to reason with such tasks, and test instructional strategies to promote the acquisition of these strategies.

We thank Cindy Hmelo-Silver for helpful comments on an earlier version of this paper.

REFERENCES

American Association for the Advancement of Science (AAAS), Project 2061. (1993). Benchmarks for science literacy. New York: Oxford University Press.

Azmitia, M., & Crowley, C. (2001). The rhythms of scientific thinking: A study of collaboration in an earthquake microworld. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings (pp. 51–81). Mahwah, NJ: Erlbaum.

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, 797–817.

Bernstein, L., Schachter, M., Winkler, A., & Wolfe, S. (1991). Concepts and challenges in life science (3rd ed.). Englewood Cliffs, NJ: Erlbaum.

Bierer, L. M., & Lien, V. F. (1984). Heath life science. Lexington, MA: Heath.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Brewer, W. F., & Mishra, P. (1998). Science. In W. Bechtel & G. Graham (Eds.), A companion to cognitive science (pp. 744–749). Oxford: Blackwell.

Bruer, J. T. (1994). Schools for thought: A science of learning in the classroom. Cambridge, MA: MIT Press.

Buchner, E. (1955). Alcoholic fermentation without yeast cells. In G. L. Mordecai & S. Fogel (Eds.), Great experiments in biology (M. G. Gabriel, Trans, pp. 27–30). Englewood Cliffs, NJ: Prentice-Hall. (Original work published 1897)

Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell & E. H. van Zee (Eds.), Inquiring into inquiry learning and teaching in science (pp. 20–46). Washington, DC: American Association for the Advancement of Science.

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

Cavalli-Sforza, V., Weiner, A. W., & Lesgold, A. M. (1994). Software support for students engaging in scientific activity and scientific controversy. Science Education, 78, 577–599.

Chinn, C. A. (1998). A critique of social constructivist explanations of knowledge change. In B. Guzzetti & C. Hynd (Eds.), Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world (pp. 77–115). Mahwah, NJ: Erlbaum.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1–49.

Chinn, C. A., & Brewer, W. F. (1996). Mental models in data interpretation. Philosophy of Science, 63 (Proceedings), S211–S219.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. Journal of Research in Science Teaching, 35, 623–654.

Chinn, C. A., & Brewer, W. F. (in press). Models of data: A theory of how people evaluate data. Cognition and Instruction.

Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically authentic scientific reasoning. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings (pp. 351–392). Mahwah, NJ: Erlbaum.

Collins, H., & Pinch, T. (1993). The golem: What everyone should know about science. Cambridge: Cambridge University Press.

Daniel, L., Ortleb, E. P., & Biggs, A. (1995). Merrill life science. New York: McGraw-Hill.

Danielson, E. W., & Denecke, E. J., Jr. (1986). Macmillan earth science. New York: Macmillan. (this is the teacher’s edition)

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

Davis, E. A., & Linn, M. C. (2000). Scaffolding students’ knowledge integration: Prompts for reflection in KIE. International Journal of Science Education, 22, 819–837.

Driscoll, M. P., Moallem, M., Dick, W., & Kirby, E. (1994). How does the textbook contribute to learning in a middle school science class? Contemporary Educational Psychology, 19, 79–100.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). Making sense of secondary science: Research into children’s ideas. London: Routledge.

Dunbar, K. (1993). Concept discovery in a scientific domain. Cognitive Science, 17, 397–434.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), The nature of insight (pp. 365–395). Cambridge, MA:MIT Press.

Dunbar, K. (1999). What scientists do and what science is: The “science as category” account of scientific thinking. Paper presented at the Cognitive Basis of Science conference, New Brunswick, NJ.

Finley, F. N., & Pocoví, M. C. (2000). Considering the scientific method of inquiry. In J. Minstrell & E. H. van Zee (Eds.), Inquiring into inquiry learning and teaching in science (pp. 47–62). Washington, DC: American Association for the Advancement of Science.

Franklin, A. (1986). The neglect of experiment. Cambridge: Cambridge University Press.

Galison, P. (1997). Image and logic: A material culture of microphysics. Chicago: University of Chicago Press.

Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. Journal of Research in Science Teaching, 33, 475–499.

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

Gitomer, D. H., & Duschl, R. A. (1995). Moving toward a portfolio culture in science education. In S. M. Glynn & R. Duit (Eds.), Learning science in the schools: Research reforming practice (pp. 299–326). Mahwah, NJ: Erlbaum.

Glaser, R., & Baxter, G. P. (1997, February). Improving the theory and practice of performance-based assessment. Paper presented at a conference of the Board of Testing and Assessment, National Research Council/National Academy of Sciences.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. Reading Research Quarterly, 28, 116–155.

Hacking, I. (1983). Representing and intervening. Cambridge: Cambridge University Press.

Hafner, R., & Stewart, J. (1995). Revising explanatory models to accommodate anomalous genetic phenomena: Problem solving in the context of discovery. Science Education, 79, 111–146.

HIRO Science Lessons (n.d.). URL: http://homepage2.nifty.com/sympathy/jikken/jikken.htm.

Hirsch, J., DeLaPaz, R., Relkin, N., Victor, J., Li, T., Karl, K., Olyarchuk, J., & Georgakakos, B. (1993). Single voxel analysis of functional magnetic resonance images (fMRI) obtained during human visual stimulation. Investigative ophthalmology and visual science, 35, 1438.

Hmelo-Silver, C. E., Nagarajan, A., & Day, R. S. (2002). “It’s harder than we thought it would be”: A comparative case study of expert–novice experimentation strategies. Science Education, 86(2), 219–243.

Hmelo, C. E., Ramakrishnan, S., Day, R. S., Shirey, W. E., Huang, Q., & Baar, J. (1998). Developing inquiry skills through scaffolded use of a simulation. In International Conference of the Learning Sciences: Proceedings of the 1998 Conference. Charlottesville, VA: Association for the Advancement of Computing in Education.

Horwitz, P., Neumann, E., & Schwartz, J. (1996). Teaching science at multiple levels: The GenScope program. Communications of the ACM, 39, 127–131.

Houghton Mifflin Interactive (1997). InventorLabs Technology. [Computer software, CD-ROM]. Somerville, MA: Houghton Mifflin Interactive.

Kim, K. (1994). Explaining scientific consensus: The case of Mendelian genetics. New York: The Guilford Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. Cognitive Science, 12, 1–48.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. Cognitive Psychology, 25, 111–146.

Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 317–330.

Knorr-Cetina, K. D. (1981). The manufacture of knowledge: An essay on the constructivist and contextual nature of science. Oxford: Pergamon Press.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. Monographs of the Society for Research in Child Development, 60 (4, Serial No. 245).

Kulm, G., Roseman, J., & Treistman, M. (1999). A benchmarks-based approach to textbook evaluation. Science Books and Films, 35, 147–153.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific fact (2nd ed.). Princeton, NJ: Princeton University Press.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. Learning and Instruction, 3, 113–132.

Linn, M. C. (2000). Designing the knowledge integration environment. International Journal of Science Education, 22, 781–796.

Loh, B., Reiser, B. J., Radinsky, J., Edelson, D. C., Gomez, L. M., & Marshall, S. (2001). Developing reflective inquiry practices: A case study of software, the teacher, and students. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings (pp. 279–323). Mahwah, NJ: Erlbaum.

Lynch, M. (1988). The externalized retina: Selection and mathematization in the visual documentation of objects on the life sciences. In M. Lynch & S. Woolgar (Eds.), Representation in scientific practice (pp. 153–186). Cambridge, MA: MIT Press.

Mallinson, G. G., Mallinson, J. B., Froschauer, L., Harris, J. A., Lewis, M. C., & Valentino, C. (1993). Science horizons (Sterling ed.). Morristown, NJ: Silver Burdett Ginn.

Maor, D., & Taylor, P. C. (1995). Teacher epistemology and scientific inquiry in computerized classroom environments. Journal of Research in Science Teaching, 32, 839–854.

McFadden, C., & Yager, R. E. (1993). SciencePlus®: Technology and society. Austin, TX: Holt, Rinehart, and Winston.

Moray, N., Lootsteen, P., & Pajak, J. (1986). Acquisition of process control skills. IEEE Transactions on Systems, Man, & Cybernetics, 16, 497–504.

Murphy, B. (1991). Experiment with light. Minneapolis, MN: Lerner.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 29, 85–95.

NASA Classroom of the Future (1996). A guide to Astronomy Village: Investigating the universe. Wheeling, WV: Wheeling Jesuit College.

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

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. Cognitive Science, 21, 109–146.

Olson, S., & Loucks-Horsley, S. (Eds.). (2000). Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Research Council.

Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. Child Development, 67, 2709–2727.

Penrose, G. (1990). Sensational science activities with Dr. Zed. New York: Simon and Schuster.

Petrosino, A. (1999). Model rockets and reflective inquiry: Design principles for effective hands-on activities. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

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

Pickering, M., & Monts, D. L. (1982). How students reconcile discordant data: A study of lab report discussions. Journal of Chemical Education, 59, 794–796.

Ramsey, W. L., Gabriel, L. A., McGuirk, J. F., Phillips, C. R., & Watenpaugh, F. M. (1986). Holt physical science. New York: Holt, Rinehart, and Winston.

Rasmussen, N. (1993). Facts, artifacts, and mesosomes: Practicing epistemology with the electron microscope. Studies in History and Philosophy of Science, 24, 227–265.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology. In S. M. Carver & D. Klahr (Eds.), Cognition and instruction: Twenty-five years of progress (pp. 263–305). Mahwah, NJ: Erlbaum.

Rosebery, A. S., Warren, B., & Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. The Journal of Learning Sciences, 2, 61–94.

Roth, W.-M. (1995). Authentic school science: Knowing and learning in open-inquiry science laboratories. Dordrecht: Kluwer.

Roth, W.-M., & Bowen, G. M. (1993). An investigation of problem framing and solving in a grade 8 open-inquiry science program. The Journal of the Learning Sciences, 3, 165–204.

Rudwick, M. J. S. (1985). The great Devonian controversy: The shaping of scientific knowledge among gentlemanly specialists. Chicago: University of Chicago Press.

Sandoval, W. A., & Reiser, B. J. (1997). Evolving explanations in high school biology. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. Journal of Experimental Child Psychology, 49, 31–57.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. Developmental Psychology, 32, 102–119.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. The Journal of the Learning Sciences, 4, 131–166.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. Journal of Research in Science Teaching, 28, 859–882.

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. Cognitive Science, 17, 285–309.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. Cognitive Science, 23, 337–370.

Schunn, C. D., & Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. Memory and Cognition, 24, 271–284.

Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding: A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. Cognition and Instruction, 15, 317–393.

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

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

Stinner, A. (1995). Science textbooks: Their present role and future form. In S. M. Glynn & R. Duit (Eds.), Learning science in the schools: Research reforming practice (pp. 275–296). Mahwah, NJ: Erlbaum.

Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific strategic support for biological inquiry. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), Intelligent tutoring systems: Third International Conference, ITS'96 (pp. 288–296). Montreal, Canada: Springer-Verlag.

Thagard, P. (1992). Conceptual revolutions. Princeton, NJ: Princeton University Press.

Thagard, P. (1999). How scientists explain disease. Princeton, NJ: Princeton University Press.

The Science House (n.d.). URL: http://www.ncsu.edu/science_house/index.html.

Theatrix Interactive. (1995). Bumptz Science Carnival [Computer software, CD-ROM]. Emeryville, CA: Theatrix Interactive.

Thompson, M., McLaughlin, C. W., Smith, R. G. (1995). Merrill physical science. New York: McGraw-Hill.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. Science Education, 86(2), 264–286.

VanCleave, J. (1997). Janice VanCleave's guide to the best science fair projects. New York: Wiley.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. Cognitive Science, 20, 75–100.

Walpole, B. (1987). Fun with science: Movement. New York: Warwick Press.

Warner, L. A., Lawson, S. A., Bierer, L. K., & Cohen, T. L. (1991). Life science: The challenge of discovery. Lexington, MA: Heath.

Whalley, M. (1992). Experiment with magnets and electricity. Minneapolis, MN: Lerner.

White, B. Y., & Frederiksen, J. R. (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.

WISE: The Web-based Integrated Science Environment (n.d). URL: http://wise.berkeley.edu/WISE/pages/research.php.

Woodward, J. (1989). Data and phenomena. Synthese, 79, 393–472.

Zimmerman, C. (2000). The development of scientific reasoning skills. Developmental Review, 20, 99–149.

Zohar, A. (1995). Reasoning about interactions between variables. Journal of Research in Science Teaching, 32, 1039–1063.