THE JOURNAL OF THE LEARNING SCIENCES, *12*(1), 5–51

Conceptual and Epistemic Aspects of Students' Scientific Explanations

William A. Sandoval
Graduate School of Education and Information Studies
University of California, Los Angeles

This article explores how students' epistemological ideas about the nature of science interact with their conceptual understanding of a particular domain, as reflected in written explanations for an event of natural selection constructed by groups of high school students through a technology-supported curriculum about evolution. Analyses intended to disentangle conceptual and epistemic aspects of explanation reveal that groups sought plausible causal accounts of observed data, and were sensitive to the need for causal coherence, while articulating explanations consistent with the theory of natural selection. Groups often failed to explicitly cite data to support key claims, however, both because of difficulty in interpreting data and because they did not seem to see explicit evidence as crucial to an explanation. These findings reveal that students have productive epistemic resources to bring to bear during inquiry, but highlight the need for an epistemic discourse around student-generated artifacts to deepen both the conceptual and epistemological understanding students may develop through inquiry.

Inquiry-based approaches to science education emphasize processes of inquiry, such as asking questions, generating and interpreting data, and forming conclusions (e.g., AAAS, 1992; National Research Council [NRC], 1996). Rarely do such formulations explicitly attend to the products of scientific inquiry. Theories, explanations, and models are examples of the kinds of artifacts of knowledge produced through inquiry. Such artifacts reflect epistemic commitments to forms of knowledge and ways of making them. Scientists do not simply form conclusions, they work to develop explanatory and predictive theories and models of the world. Understanding how to conduct scientific inquiry thus relies on an understanding of the goals for that inquiry, an understanding that includes the forms of knowledge de-

---

Correspondence and requests for reprints should be sent to William A. Sandoval, 2339 Moore Hall, Box 351521 Graduate School of Education and Information Studies, University of California, Los Angeles, Los Angeles, CA 90095–1521. E-mail: sandoval@ucla.edu

sired as products. The goals for these products ground scientists' processes of inquiry. For example, the canonical experimentation strategy of controlling variables is valued because it allows for the isolation of causal relations. If that were not the goal, there would be no reason to control variables. This point may seem obvious to a scientist, but children do not always see the goal of experimentation, at least in school, as isolating causal relations (Reif & Larkin, 1991; Schauble, Glaser, Duschl, Schulze, & John, 1995). Support for science inquiry, then, needs to help students do more than understand the processes of inquiry, it has to help them understand the point. Beyond this, isolated causal relations are not the stuff of scientific knowledge, but are woven into broader, coherent explanations of the world. Helping students to see the formal features of these products of inquiry should aid their own efforts to construct scientific knowledge.

In this article I examine what happens when students are explicitly supported in understanding the kind of product their inquiry should produce. I explore how students' ideas about the nature of science, their epistemological ideas about science, interact with their conceptual understanding of a particular domain, as manifested in the artifacts they create from inquiry. These artifacts, written explanations about an event of natural selection, were constructed by groups of high school students through a technology-supported curriculum about natural selection and evolution. The broad effort of this project, the Biology Guided Inquiry Learning Environments (BGuILE), has been to understand how to combine both strategic and conceptual scaffolds for students' inquiry in this domain (Reiser et al., 2001). The approach is to scaffold students' into the epistemic game (Collins & Ferguson, 1993) of scientific explanation with a software program, ExplanationConstructor (Sandoval, 1998), that integrates domain-specific guidance about what to explain in a particular problem with guidance about what a good scientific explanation looks like. This epistemic guidance focuses on two criteria for explanations: (a) the coherent articulation of causal claims and (b) using evidence to support or refute claims.

In the first part of this article I argue that epistemic guidance for inquiry must be integrated with conceptual guidance for particular domains, and I describe ExplanationConstructor and aspects of the curriculum in which it is embedded that are designed to provide such guidance. Briefly, there are at least three reasons that inquiry-based reform approaches to science education need to attend to epistemic issues. First, epistemological development is an explicit goal of the reforms (e.g., AAAS, 1992; NRC, 1996). Second, as I describe in the following, there are good reasons to believe that students' ideas about the nature of science influence their efforts to conduct science. The third reason is methodological. Namely, the widespread finding that students hold naive epistemological beliefs about science (see reviews by Abd-El-Khalick & Lederman, 2000; Lederman, 1992) has been generated largely without examining students' own scientific practices. In part, the effort to make explicit the epistemic goals of inquiry is an effort to develop a better understanding of the epistemological ideas that students are able to bring to bear on their inquiry efforts. After discussing these issues, I describe students' use of ExplanationConstructor to write explanations for a historical example of natural selection in the Galápagos islands, as part of a month long unit on evolution collaboratively designed by our research team and a high school teacher.

In the second part of the article I present analyses of the explanations that students wrote for this Galápagos problem. These analyses attempt to answer three questions. First, can students use the integrated conceptual and epistemic guidance of ExplanationConstructor to articulate explanations consistent with the theory of natural selection? Second, can students meet the previous criteria for causal coherence and evidentiary support in their explanations? These two questions concern what I distinguish here as conceptual and epistemic understanding, respectively. If students can write natural selection explanations, that suggests they understand the theory and how to apply it to answer particular problems. If students can write coherent and well-supported explanations it suggests they understand the epistemic game of explanation. The third question for this study is, how do these two aspects of understanding, conceptual and epistemic, interact during inquiry? There are good reasons, discussed later, to view scientific epistemologies as grounded within specific disciplines. An epistemic understanding of the criteria for explanations ought to help guide students' work, but in itself is not enough to enable students to be able to explain a particular problem. Students need to understand the disciplinary concepts and theoretical frameworks that could explain a particular problem. Of course, inquiry-based instruction is intended to help students learn such disciplinary concepts. Disciplinary scaffolds grounded within explicitly epistemic structures might guide students' inquiry and help them to see how to use disciplinary concepts to explain particular events.

My goal in this article is not to show how much students learn from their use of ExplanationConstructor. Instead, my goal here is to uncover some of the underlying epistemic ideas students show through their explanations of a complex event. This study contributes to an emerging body of research suggesting the situated character of students' ideas about science (e.g., Hammer, 1994; Roth & Roychoudhury, 1994), and illuminates how those play out during their inquiry. I will attempt to make clear how features of the design of ExplanationConstructor and the enactment of the curriculum supported and possibly hindered students' efforts.

SCIENCE AND SCIENCE LEARNING AS SITUATED, EPISTEMIC PRACTICES

The design of ExplanationConstructor and the curriculum described in this article rely on situated theories of learning as apprenticeship (Lave & Wenger, 1991; Rogoff, 1990), and especially the model of cognitive apprenticeship (Collins, Brown, & Newman, 1989). From this perspective, learning science entails the appropriation of discipline-specific modes of discourse and action. These ways of talking, thinking, and acting include often tacit epistemological commitments, commitments to the kinds of questions worth asking, the kinds of answers worth having, and acceptable methods for making them. Developing an apprentice-oriented science pedagogy thus requires an epistemic focus, an effort to understand how knowledge is made within a discipline. In this section, I summarize recent trends in the philosophy of science to connect epistemology to methods of inquiry in science. I also briefly review what cognitive research has to say about students' scientific epistemologies, and note that it has generally ignored or played down disciplinary differences. Together, these literatures underlie my approach to integrating conceptual and epistemic scaffolds for explanation, described in the next section.

First, some clarification of terms. The term *epistemology* is used by different authors to mean very different things (for a review, see Hofer & Pintrich, 1997). I use the term throughout this article to refer to epistemologies of science: beliefs about the nature of science and scientific knowledge. This is different from, albeit related to, students' beliefs about themselves as learners of science, or their beliefs about how science is learned (Hammer, 1994; Hogan, 2000). In the sense I use here, students' epistemologies of science include their ideas about what scientific theories and explanations are, how they are generated, and how they are evaluated as knowledge claims.

Disciplinary Epistemologies of Science

Epistemology has long been a central topic in the philosophy of science. In particular, over the last 50 years, debates have raged over the nature of scientific knowledge, its relation to truth, its correspondence to reality, and so on (an excellent summary of these debates is given by Driver, Leach, Millar, & Scott, 1996). These debates make clear that there is no single consensus epistemology of science. At the same time, there are broad epistemological commitments probably shared by all of the natural sciences, at least. These include the goal for causal explanations for events in the world, that such explanations be parsimonious, and that they account for observations. They also include the notion that explanations are social and historical constructions, rely on scientists' creativity, are tentative, and that theories are generally not abandoned until a better one comes along. These are all aspects of the kinds of products—the intellectual artifacts—that scientists endeavor to construct through their inquiry.

Considerations of the nature of scientific knowledge go hand in hand with ideas about the means for generating scientific knowledge. Again, speaking broadly, all of the sciences share a commitment to empirical data and to systematic comparison of observations. This is because the goal for causal explanations requires that alternative causes be ruled out, and that requires systematic comparison. Thus, the goals for the desired product drive the methods (i.e., processes) used to construct them. The canonical form of such comparison is controlled experimentation, where potential causal relations between candidate variables are tested through systematic variations in one variable at a time. Not coincidentally, experimentation has been the primary focus of cognitive studies of children's scientific thinking (summarized in the following).

Yet, there are many scientific disciplines in which experimentation is not the norm, or even possible. One such discipline is evolutionary biology, where in most cases experimentation on natural populations is not possible. Instead, field biologists systematically observe populations in their environments, often over long periods of time, and look for relations between factors in the environment and features of organisms. Such differences in methods are tied to differences in the kinds of questions disciplines attempt to answer, and they lead to differences in theories. Evolutionary theory, for example, can provide historical explanations but cannot generate precise predictions about the evolution of particular populations over time (Mayr, 1988). Thus, questions, theories, and methods vary across disciplines. Within disciplines, research paradigms also shift historically, and such shifts change the questions asked, the answers deemed valuable, and the acceptable methods for generating them (Kuhn, 1970).

For science education there are at least two implications from this. One is that part of what it ought to mean to know a discipline should include some knowledge of the kinds of questions and theories that the discipline pursues. That evolutionary theory is not predictive is an important facet of biology, and different from, for example, physics. The second implication is that students need to understand the goals for the products that their inquiry processes are intended to produce. In the broad sense, this means that students should learn that scientific explanations are efforts to construct causal accounts for how or why things happen, and that they must account for observations. Within a discipline, such as evolutionary biology, students should also learn what kinds of causes make sense within the discipline. This entails more than just being able to recite Darwin's theory, but to be able to use it to explain actual events.

Students' Epistemologies of Science

Most students do not seem to have an epistemology of science that is consistent with current inquiry-based approaches to learning science. Few students see science as a process of building and testing models and theories; instead, science is seen as a steady accumulation of facts about the world (Carey & Smith, 1993; Driver et al., 1996; Lederman, 1992; Linn & Songer, 1993). Many students do not distinguish experimental findings from the ideas they are designed to test, or see that relation simplistically: Experiments tell you straightforwardly if you are right or wrong (Carey, Evans, Honda, Jay, & Unger, 1989). Students often do not see that experiments are intended to test causal relations (Reif & Larkin, 1991; Schauble et al., 1995). Thus, students ideas about the kinds of products scientists produce hinders their understanding of scientific processes.

When students are provided explicit explanatory goals with which to explore domains, however, they conduct more effective experiments. Dunbar (1993), for example, found that subjects who were asked to explain data, rather than verify a given hypothesis, were more systematic and designed better experiments and were thus more likely to discover the correct function of a gene. Schauble and her colleagues (Schauble et al., 1995) similarly found that 5th grade students could design better experiments after explicit instruction that experiments are intended to isolate causal relations. These results suggest that making the epistemic demands of inquiry explicit to students can improve their efforts.

These studies and others on students' scientific thinking (Klahr, Dunbar, & Fay, 1990; Klahr, Fay, & Dunbar, 1993; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Schauble, & Garcia-Mila, 1992; Schauble, Glaser, Raghavan, & Reiner, 1991; Schauble, Klopfer, & Raghavan, 1991; Shute, Glaser, & Raghavan, 1989) have almost exclusively looked at controlled experimentation. Students' inquiry into the Galápagos problem described in the following was not experimental, but historical as they attempted to construct an explanation for an event that had already happened. Many of the issues, however, are the same as students had to decide what data to look at, what comparisons to make, and how to interpret them. This study differs from that previous work in an important way. These previous studies have either asked students to verify a singular causal claim or to infer which of a set of variables is a causal factor in some phenomenon. Here, students were asked to construct chains of causal inferences into a coherent explanation. This explanation task introduces an epistemic component absent in these other studies, as students have to decide for themselves the level of causal specificity for their explanations. Students could understand what happens, and why, but leave out parts of their explanation that they do not think are important, or are obvious, and so on.

Another important difference between this study and prior studies on both experimentation strategies and students' scientific epistemologies is that here I am most interested in understanding students' ideas about the nature of science as they manifest themselves in their actual practice of inquiry. Surprisingly, given the interest in students' professed epistemological beliefs about science, comparatively little work has examined how such beliefs influence students' efforts to learn science. Students' professed strategies for learning science seem generally consistent with their professed conceptions about the nature of science (Hammer, 1994; Songer & Linn, 1991), although students' views of themselves as science learners may be more predictive than expressed ideas about formal science (Hogan, 1999).

An important finding from recent work is that students with more sophisticated epistemologies seem to take better advantage of inquiry-based learning opportunities: constructivist beliefs seem to encourage constructivist learning strategies (Linn & Songer, 1993; Windschitl & Andre, 1998). A crucial gap in these studies, however, is that epistemological conceptions are assessed independently of students' performance on inquiry activities. Therefore, whereas an important relation between epistemological understanding and inquiry appears to exist, there is little knowledge about how epistemological conceptions affect students' reasoning during inquiry, or even what epistemological ideas students use to construct their own scientific understanding. By making explicit the goals for students' inquiry and supporting their efforts to construct explanations, as described next, this study seeks to illuminate students' own epistemic practices as reflected in artifacts generated from their own inquiry.

The epistemic demands of inquiry include the criteria to which knowledge claims are held. There are many criteria to which we may want to hold scientific theories and explanations (cf. Gitomer & Duschl, 1995). In this work I have primarily focused on two epistemic criteria for students' explanations: causal coherence and evidentiary support. The criterion for causal coherence actually embodies two epistemic goals for scientific explanations: (a) that they articulate causal mechanisms to explain phenomena, and (b) that chains of causes and their effects cohere sensibly. The criterion for evidentiary support reflects the idea that explanations are constructed to explain patterns of data, and so it should be clear how data relate to claims. Although there are certainly other useful criteria that we might wish students to keep in mind during inquiry, these two are central to the notion of scientific explanation (Kuhn, 1970; Mayr, 1988). The scaffolds for explanation in ExplanationConstructor are designed to support students in meeting these criteria, while grounding them in conceptual guidance for specific domains.

Students' Conceptual Understanding of Natural Selection

The theory of evolution through natural selection is a unifying theory of modern biology, and notoriously difficult for students to understand. One difficulty is that students tend to think "typologically," they see individuals as representative of an entire population (Greene, 1990). Such a view makes it difficult to see the importance of individual variation, which is crucial to understanding the explanatory power of the theory. Also, students fail to recognize that variations must already exist to be selected (Bishop & Anderson, 1990; Brumby, 1984; Settlage, 1994). Instead, many students believe that individual organisms change their traits in response to environmental pressures (e.g., giraffes grew long necks because they had to reach leaves high up in trees). Another source of difficulty is that students do not understand the mechanism of inheritance through which traits can be passed on, often believing that acquired characteristics can be inherited (Wood-Robinson, 1995). That is, although selection operates on specific individuals within a population, through variations in traits, its effects are seen in populations, as the distribution of traits (i.e., individuals with certain traits) within a population changes over time.

A further potential difficulty for students learning the theory of natural selection is that the theory itself is so broad, it “explains nothing, because it explains everything” (Lewontin, quoted in Mayr, 1988, p. 97). That is, the particulars of what might constitute a selective pressure or an advantageous trait are essentially unique, dependent on specific organisms in specific environments. It is therefore impractical, even unreasonable, to expect that students could induce the theory of natural selection by examining certain cases. The approach taken in BGuILE is to focus students’ inquiry on theory articulation (Ohlsson, 1992) rather than theory building (see Reiser et al., 2001). Given explanatory frameworks, can students use them to explain particular events?

INTEGRATED CONCEPTUAL AND EPISTEMIC SUPPORTS FOR EXPLANATION

Recent technologies designed to support inquiry include tools to structure students’ representation of their own thinking about scientific phenomena. These tools provide general structures for the articulation of arguments (Bell & Linn, 2000; Suthers, Weiner, Connelly, & Paolucci, 1995) or building models (Jackson, Stratford, Krajcik, & Soloway, 1994). ExplanationConstructor was designed to focus students’ inquiry on the epistemic goals for the creation and evaluation of causal explanations, with general scaffolds couched within domain-specific conceptual frameworks.

ExplanationConstructor was designed as part of a suite of software tools that form the core of an inquiry-based unit on evolution (Reiser et al., 2001; Tabak, Smith, Sandoval, & Reiser, 1996). In the following, I describe the context of my study and the outline of the evolution unit as enacted in this study. The description focuses primarily on students’ activity during an investigation of a problem of natural selection among finches on a small Galápagos island. Following this description of the context, I describe the design of scaffolds built into ExplanationConstructor. My intent in these descriptions is to explain how the approach to integrating epistemic and conceptual scaffolds is embodied within the software and curriculum, and suggest how aspects of both the software and the classroom context contributed to students’ explanation of this complex problem.

Classroom Context of Use

Setting

The BGuILE evolution unit was used in three introductory high school biology classes, taught by one teacher, during the spring term of 1997. This high school is in an affluent suburb of a major Midwestern city. At the time of this study, the school was approximately 85% White, with less than 2% of students on a school lunch program. There were 69 students in the three classes, with 24 in an honors class. Students were admitted into the honors class on the basis of a written placement test unrelated to biology. Thus, these students may have been more motivated to take an honors class, and possibly better writers, but pretests showed no differences between classes on prior knowledge of evolution (Tabak, 1999). The teacher reported that his honors class occasionally wrote essays, but the other two classes did not do much writing. I point this out because the explanatory task described here was writing intensive, thus the quality of explanations could differ between classes. This evolution unit began 5 weeks into the second term of the year.

The evolution unit occurred over 4 weeks, as summarized in Table 1. The focus of the analyses I present here are the explanations that student groups constructed during their investigation of a case of natural selection among a species of finch on a Galápagos island (Grant, 1986), during the 2nd week of the unit. This investigation was the first of two complex computer-based inquiries that students conducted during this unit. By the time students began working on the finch problem they had some introductory knowledge of Darwin's theory of natural selection. Also, during the 1st week of the unit students conducted a variety of classroom labs concerning individual variations within populations, and the effects of individual trait variations in different environments (Tabak & Reiser, 1997). They also, as described by Tabak and Reiser, completed an introductory investigation about marine iguanas that introduced them to many of the forms of data representation they would see in the finch investigation, to explanation guides, and to the relations between trait differences and functions they enabled. For brevity, I do not describe the unit as a whole in detail here, focusing instead on students' activity during the finch investigation.

I have chosen to focus on students' explanations for the finch problem primarily because they represent their first attempts to apply the theory of natural selection to explain a particular event. They can thus show how well students can use the combined conceptual and epistemic scaffolds to construct explanations of natural selection. They also reflect students' understanding of the epistemic goals for scientific explanations. This is particularly important in relation to previous studies of students' scientific epistemologies, as those studies have not yet examined how students' epistemic views are manifested through their own inquiry.

TABLE 1
Main Activities of Each Week of Evolution Unit

Week	*Major Activities*	*Goals*
1	"Darwin" visits class, describes theory, takes questions from students.	Introduce major ideas of theory, and historical context when it appeared.
	Labs to measure variation in plants (sizes of peas) and humans (hand spans).	Emphasize normal distribution of traits, and individual differences; introduce graphical representations used in investigations.
	Paper-and-pencil iguana investigation: why two groups of same species of iguana live in different locations on a small island.	Introduce Galápagos ecosystem, and focus on individual differences and structure–function relations.
	Discussion of explanations.	Elicit students' criteria for explanations.
2	Computer-based Galápagos finch investigation.	Apply theory of natural selection to explain a real event.
3	Cartoon movie of evolution.	Visualize effects of random mutation, natural selection, and sexual selection on evolution.
	Human evolution labs: sequence hominid skulls; compare selected traits of humans and gorillas.	Explore pathway of human evolution; similarity to primates.
	Begin TBLab—second computer-based investigation.	Emphasize role of pre-existing variation in natural selection; second chance to apply theory to explain complex event.
4	Complete TBLab.	Completed TB investigation. Ended unit with class discussions centered on mapping finch and TB investigations into theory of natural selection.
	Unit wrap-up.	Whole-class discussions centered on mapping finch, colored dots, and TB investigations into theory of natural selection.

The Finch Investigation

The finch scenario poses the following problem to students: During the year 1977 the population of medium ground finches on the tiny Galápagos island of Daphne Major fell by more than 60%. Why did so many finches die during 1977? More importantly, why were the surviving finches able to survive? Students used a computer-based investigation environment (Tabak & Reiser, 1997) to examine several sources of data that could be used to construct possible explanations to answer these two questions. These data included rainfall amounts, seed types and amounts, finch predator data, and several kinds of physical (e.g., weight, beak length) and behavioral (e.g., foraging, mating) data about the ground finches. By design, there was no single piece of data that students might look at that would generate a solution. Instead, students had to coordinate multiple sources of data to understand what happened to the finches and why some could survive when most could not.

This event is an important case of documented natural selection in the wild (Grant, 1986; popularized in Weiner, 1994). Grant's explanation for what happened to the ground finches on Daphne Major during 1977 is that the wet season of 1977 essentially never happened, causing a severe and prolonged drought. This drought in turn drastically diminished the supply of seeds these finches rely on for food. Not only were the amount of seeds greatly diminished, but the type of seeds drastically altered. The soft seeds preferred by the finches were quickly eaten up, leaving only a hard-shelled seed known as tribulus. Tribulus seeds are covered by a hard, spiked shell that encloses four to five seeds that the finches can eat. Only the finches with larger than average beaks were able to open tribulus shells. By the end of 1977, only the largest-beaked birds in this population had survived. When they mated during the following wet season, their offspring tended to inherit larger beaks; consequently the average beak size in the population increased. That is natural selection in a nutshell.

Framing the Problem

The finch problem was posed to students as the two questions previously mentioned. On the first day of the unit, a week prior to the finch investigation, the teacher visited each of his three classes impersonating Darwin. He briefly described Darwin's theory of natural selection: Because more individuals are born into a population than can survive, and because individuals vary, some individuals are more likely to survive than others. Because individual traits are inherited by offspring, the result of this selection will be changes in populations over time. The discussion of the theory on this day was quite brief, no more than 10 min, with the remainder of the 40-min period spent discussing the religious and social climate of Darwin's England. The teacher also used this time as Darwin to indicate the ideas Darwin had relied on to formulate his theory. He explicitly mentioned the influence on Darwin's thinking of Malthus' economic model of overpopulation and Lyell's geology.

The variation labs that students conducted during the next few classes were intended to emphasize the amount of individual variation within populations, as most students overlook its importance (Brumby, 1984; Greene, 1990; Settlage, 1994). Students had no explicit instruction on the habitat of the Galápagos islands, except for their experience with the iguana investigation, or on the physiology or ecology of birds. Instead, students were expected to learn important features of the island environment and of the birds from their exploration of data within the computer environment.

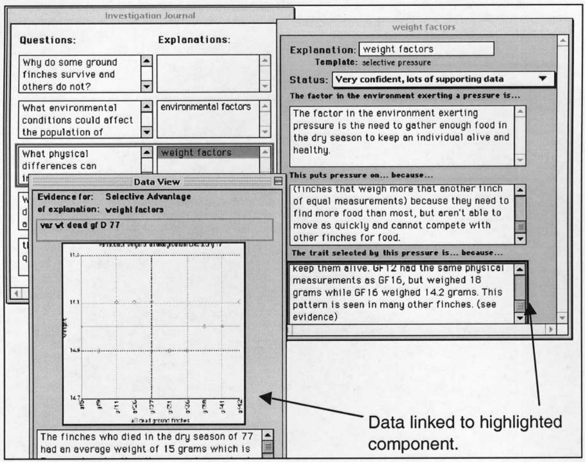
The day before the start of the investigation, the teacher held a discussion in each class in which students were asked how they decided whether an explanation was any good. Prior to this discussion, the teacher and researchers decided that they would not impose a definition of a good explanation on students, but instead prompt students to consider their own criteria. Students suggested that explanations had to be backed up by relevant evidence and they had to be causally coherent, consistent with research showing adolescents' sensitivity to causal mechanism (Koslowski, 1996). Students raised several issues relating to the breadth and relevance of the evidence used to back up an explanation. Specifically, many students spoke of the need for an appropriate and large enough sample of data to answer a question. A typical example of the way students explained this demand was a suggestion from one student that to study how many students in the school wore makeup, one would have to survey a large number of both boys and girls, surveying only girls or boys would skew the results. Many students in each of the three classes also mentioned that explanations had to "make sense," although they were less clear about how that should be judged. One or two students in each class mentioned the idea that an explanation had to be "like a chain" of causes and effects, or that claims "have to follow after each other." Students ideas were written down on the board during discussion, but particular elements in this emergent rubric were not publicly evaluated or validated by the teacher. Although various criteria were voiced in each class, I do not know the extent to which these criteria were shared by all students.

As students, working in groups, were introduced to the problem, the researchers, including myself, walked them through a demonstration of the software, both the finch investigation environment and ExplanationConstructor. This demonstration served several purposes. First, students learned how to use the software, and they became aware of the scope of data available to them. Second, they were given an idea of the nature of the work product, explanations, that they were expected to produce. Students were encouraged to rephrase the driving investigative questions in their own words and to record subsidiary questions that they thought might be important to answer along the way to answering the driving questions. Following the software demonstration, the teacher explicitly framed students' task as explaining *how* the birds died and *why* other birds survived. He told students to make at least two explanations and told them that their grade would depend on how well they used data to support their claims. There was no discussion in any of the classes about how the use of data might be judged.

Investigation and Explanation

Students worked on the finch investigation in groups of three per computer for 5 days, roughly 4.5 hr of instruction. ExplanationConstructor and the finch investigation environment ran jointly such that students could freely move between them. The finch environment was the place where students could query data about the birds and their island habitat. ExplanationConstructor was the place where students wrote their investigation journal, recording the questions they were trying to answer and their current explanations. Figure 1 shows a sample journal from one group (from the end of the investigation), including the major components of ExplanationConstructor. When students first opened their journals they were given a window where they could record their questions (upper left in Figure 1). Students were encouraged during the software demonstration to record the overarching questions here.

Creating explanations. Groups were free to examine data or record questions and explanations in any order they liked, thus the problem was quite open-ended. When ready to create an explanation, groups first selected an *explanation guide*, a specific guiding framework for their explanation. During the software demonstration, students were instructed that they needed to decide what kind of explanation they were making, and that different guides could explain different problems. Each available guide had a brief summary description of its purpose. Once a guide was chosen, groups were given a new window with a blank template. Groups named explanations so that they could return to work on them later (e.g., "weight factors" in Figure 1).



Questions:

Why do some ground finches survive and others do not?

What environmental conditions could affect the population of

What physical differences can

Explanations:

environmental factors

weight factors

Evidence for: Selective Advantage
of explanation: weight factors

var wt dead gf D 77

The finches who died in the dry season of 77 had an average weight of 15 grams which is

Explanation: weight factors
Template: selective pressure
Status: Very confident, lots of supporting data

The factor in the environment exerting a pressure is...

The factor in the environment exerting pressure is the need to gather enough food in the dry season to keep an individual alive and healthy.

This puts pressure on... because...

(finches that weigh more that another finch of equal measurements) because they need to find more food than most, but aren't able to move as quickly and cannot compete with other finches for food.

The trait selected by this pressure is... because...

keep them alive. GF12 had the same physical measurements as GF16, but weighed 18 grams while GF16 weighed 14.2 grams. This pattern is seen in many other finches. (see evidence)

Data linked to highlighted component.

FIGURE 1 ExplanationConstructor1.0, used in this study, showing one group's journal. The "Investigation Journal" window, upper left, is where students recorded their questions and organized explanations for them. The window at right shows a group's survival explanation. At the lower left is a piece of data linked to the last component of the explanation, highlighted with a bold border.

ExplanationConstructor: Integrated Epistemic and Conceptual Scaffolds

ExplanationConstructor is an electronic journal where students record the questions they are trying to answer and construct candidate explanations for these questions, as they explore computer-based investigation environments. The primary form of support in the software are explanation guides. There are three key features of explanation guides intended to support students' construction of causally coherent, well-supported explanations. First, each explanation guide represented a specific theoretical framework relating to evolution or ecosystems. Second, every explanation guide represented a similar general causal structure for explanations. Third, ExplanationConstructor provided a facility to link specific pieces of data from the investigation environment to particular causal components of students' explanations.

Theories as Explanatory Frameworks

Scientific theories are generative explanatory frameworks that provide a way to make sense of particular phenomena. Components of a particular theory frame the way a particular phenomenon might be explained. Explanation guides provide a conceptual framework to help students decide what needs to be explained about the current problem being investigated. For instance, the "selective pressure" guide shown in Figure 1 maps the major components of a natural selection explanation that students need to articulate. First, students are asked to identify the factor in the environment exerting a pressure (e.g., the drought). Then they are asked to identify who is pressured by this (e.g., the finches) and why (e.g., lack of food). Finally, the guide indicates that the explanation should identify the trait selected (e.g., longer beaks) and the reason why (e.g., because they enable finches to crack tribulus shells). The particular wording of guide prompts and their sequence was derived through a combination of consultation with professional biologists and pilot testing with adolescents prior to this study. The set of guides available for this problem in this study are shown in Table 2.

Different groups used these explanation guides in different ways, and the same group's use of guides varied at different times in their investigations. Sometimes groups created explanations when they thought they had reached some conclusion. Other times, groups created new explanations simply to keep track of data that they decided was important, although this was less common. Once an explanation was created some groups would fill out one component and then return to their exploration of the data. These subsequent explorations were sometimes explicitly driven by a group's interest in completing the next component. Other times groups filled in entire explanations and then returned to the investigation environment to

TABLE 2
Explanation Guides Available During Finch Investigation

Name	*Description*[a]	*Prompts*[b]
Character divergence	Competition for resources (such as food) within a geographic area can lead species to diverge toward the extreme variations of a trait or set of traits. Individuals will tend to group within areas where, due to some characteristics they have, they can lessen competition for resources. This divergence sometimes leads to new species.	The resources over which individuals compete are … The location of these resources varies … Use of these resources by individuals varies with the following trait(s)… Individuals diverge to one of the locations because…
Environmental catastrophe	An environmental catastrophe is any event which can cause a lot of damage to an environment (like a hurricane or an oil spill). Whether or not an individual organism is killed or otherwise affected by such an event is usually due to bad luck rather than some characteristic of the individual.	A catastrophic event occurred … Individual … were affected because … The overall effect on the population was …
Predator–Prey	Relations between predator and prey in an ecosystem are usually balanced according to some set of factors. When those factors change, it may change the relationship. For example, an unusually large population of rabbits may make the population of foxes increase dramatically.	The normal predator-prey relation is … The factor in the relationship that has changed is … The effect of this change on the relationship is …
Selective Pressure	A selection pressure is some factor in the environment which cause some trait(s) in an organism to be selected for. Organisms with these trait survive better, or have more chances to have offspring, than those without. Sometimes a drastic change in the environment can produce a selection pressure.	The factor in the environment exerting a pressure is … This puts pressure on … Because … The trait selected by this pressure is …
Intraspecies niches	Groups within a single species can occupy different niches within their environment due to differences in structure and/or behavior. Occupying such niches can often lessen competition for needed resources, like food or potential mates.	Environmental differences between areas are … These differences pose the following threats or benefits … Individuals from one area share these physical characteristics … different from individuals from another area … The advantage to … of being separate from … is …

[a] Descriptions shown in guide selection dialog. [b] Prompts labeled explanation components.

collect supporting evidence. The specific process for constructing explanations was not constrained by the software or by the teacher. Instead, the teacher and researchers stressed that groups should thoroughly explore the available data, consider multiple hypotheses, and use data to support their claims.

Causal Structure

While representing domain-specific explanatory frameworks, ExplanationConstructor also highlighted the general structure of causal explanations as chains of causes and consequent effects. Explanation guides visually represented a sequence of separate components of explanations. Explanation components were marked by prompts that rhetorically and conceptually joined components together (Figure 1). These prompts were given in domain-specific terms, but they functioned in a more general way to emphasize that there are separate components to explanations that have to hang together in a coherent manner. In this sense, explanation guides are an *epistemic form*, a particular knowledge representation that affords particular *epistemic games*, reasoning strategies and manipulations of the representation that allow particular forms of knowledge construction (Collins & Ferguson, 1993). I expected that the epistemic form of explanation guides would help students to play the epistemic game of constructing coherent, well-supported causal explanations, articulated in terms of the relevant domain theory of natural selection.

Linking Data to Claims

A crucial aspect of ExplanationConstructor was to represent data as distinct from, but linked to, the causal claims within an explanation. In part, this was an effort to help students disentangle potential confusion between explanations and their evidence (cf. Kuhn et al., 1988). With ExplanationConstructor students linked particular data to specific causal claims and justified the relevance of that data as evidence. In the version of the program used in this study, students first selected the causal component they wanted to link evidence to (the component with the bold border in Figure 1), and then pasted their data into a list of data linked to that component. The separate representation of claims and data encouraged students to think about the causal claims they were trying to make and the data that supported them. As students generated these data through their investigations, they actively selected those data they believed supported specific claims.

A Vignette: Evan, Franny, and Janie

A brief vignette of one group, a boy, Evan, and two girls, Franny and Janie, may provide a sense of how students worked through this problem. This is meant only to illustrate what the experience was like for students, how they interacted with each other as they explored the problem and used ExplanationConstructor to record their thinking and make sense of the problem.

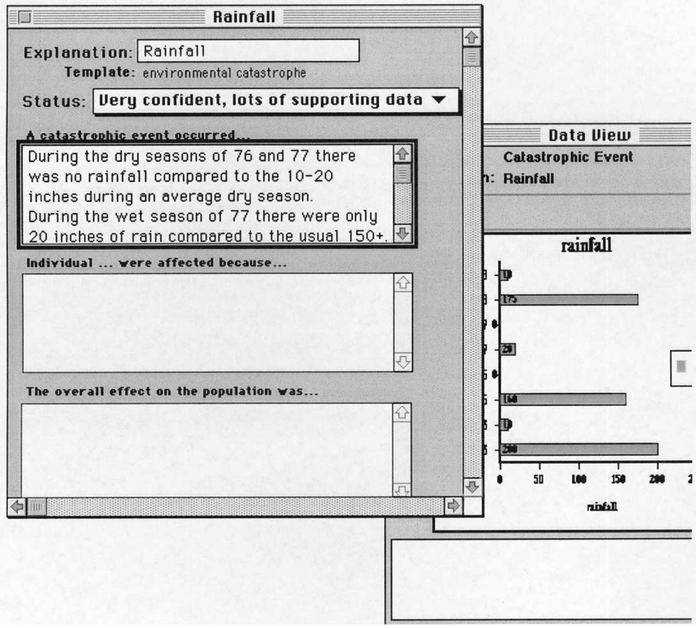
Evan, Franny, and Janie began their investigation of this problem thinking that their goal was merely to explore the environment and find out something interesting about the finches. They spent the first 20 min of their investigation looking at mating patterns across years. They eventually became frustrated and asked a researcher to help them clarify their observations about mating. Instead, the researcher refocused their efforts on the two driving questions. The group immediately looked at the weather data available in the finch investigation software and noticed the lack of rain in 1977. They agreed this was important and decided to create an explanation to record their finding. They selected the environmental catastrophe guide (Table 2) and summarized their interpretation of the rainfall graph they had found, as in Figure 2.

They then had the following exchange about where to go next (throughout their collaboration, Janie was at the keyboard).

Janie: So, now where do we want to go?

Franny: You guys, we need sub-questions.

Evan: No, we're still answering that question [pointing to their journal].



Explanation: Rainfall
Template: environmental catastrophe
Status: Very confident, lots of supporting data

A catastrophic event occurred...
During the dry seasons of 76 and 77 there was no rainfall compared to the 10-20 inches during an average dry season.
During the wet season of 77 there were only 20 inches of rain compared to the usual 150+.

Individual ... were affected because...

The overall effect on the population was...

n: Rainfall

FIGURE 2 One group's initial explanation for why the finches died.

Franny: We are?
Evan: Yeah.
Janie: We are?
Evan: Yeah, we have to find out which individuals were affected [pointing to prompt for second component of explanation in Figure 2].

Here, and several other times throughout their 5 day investigation of this problem, the group used the prompts to keep track of where they were at in their investigation, what questions they had asked and answered, whether they had data for their explanations, and so on. Notice that the group is not simply monitoring their progress here, but doing so in epistemically important terms. Their attention is focused on determining which epistemic entity, a question or an explanation, they need to construct.

Detailed observations were collected on only a small number of groups, but field notes corroborate that most groups' talk frequently centered around the emergent ideas they were recording in their journals and where the prompts suggested they should go next. I should point out here that as the investigation wore on some of the important ideas that Franny, Janie, and Evan articulated about the finches were not recorded in their journals. I consider the implications of this in the discussion. For the moment, I hope this vignette provides some sense of students' collaborative exploration of this problem.

As for the teacher, he rarely appeared in our videotaped records, although field notes indicate that he was more or less continually cycling among groups. The researchers were also visiting groups, although our role was primarily to help students with technical problems or explain how to get some particular data. The teacher, from our observations, primarily asked students to report their progress and to describe the data they were using to generate specific claims. I do not know how much he encouraged students to specifically cite data in their explanations.

Groups in each of the three classes worked independently of each other. There was, at least during our classroom observations, little talk across groups during their investigations. At the end of the last period allotted to this investigation, students saved their journals in ExplanationConstructor. These were collected by the research team, and printed versions of journals were returned to students to hand in to the teacher.

EXPLANATIONS AS ARTIFACTS OF STUDENTS' CONCEPTUAL AND EPISTEMIC UNDERSTANDING

Selection of Explanations

Twenty-two groups of students from the three classes investigated the finch problem over the course of 1 week. All of the questions they recorded, the explanations they created, and the data they linked as evidence were captured in their ExplanationConstructor journals. The journals of 19 groups were successfully retrieved (three groups' journal files were lost due to problems with the school's computer network).

Recall that students were asked to solve two interrelated questions: How did so many finches die? Why did some finches survive? Groups took varying paths to explaining these questions, including asking different questions along the way and recording many different explanations. Many of the explanations in groups' journals were intermediate; ideas briefly pursued by the group and then abandoned. Initially, I intended to analyze groups' explanations to each of the two posed questions. Some groups, however, perceptively noticed that the two questions as posed are really two aspects of the same question: The cause of the finches' death is the selection event. This, combined with the variation in the number of explanations groups constructed, focused my analyses on students' explanations for the finches survival. With regard to this domain, the question of survival is the key question of natural selection. The most complete explanation offered by a group was assumed to be their intended best explanation. In the few cases where two fairly complete explanations were written by a single group, both explanations were selected for initial analysis.

Methods of Analysis

Three questions framed this analysis. First, can students use the integrated conceptual and epistemic guidance of ExplanationConstructor to articulate explanations consistent with the theory of natural selection? Second, can students meet criteria for causal coherence and evidentiary support in their explanations? Third, how do these two aspects of understanding, conceptual and epistemic, interact? Answering these questions requires analyses that can separate the epistemic aspects of explanation from the conceptual understanding of the specific problem as reflected in groups' written work. I have tried to approach the problem in two ways, related to the two epistemic criteria of interest: causal coherence and evidentiary support. First, I developed an assessment of the overall quality of groups' explanations, primarily to answer the first question of whether or not students could successfully articulate natural selection explanations. Second, I developed an assessment of the causal coherence of explanations that is independent of their biological accuracy, as a direct assessment of groups' abilities to satisfy the criterion for causal coherence. I also examined groups' explicit citations of data in their explanations, as a measure of their understanding of the need to explicitly support claims with data.

Overall Quality

I consider the overall quality of students' explanations for this problem as the extent to which students are able to (a) *articulate* causal claims within a natural selection framework, and (b) whether their claims are *warranted* from the data they examined during investigation. For this purpose, a good natural selection explanation articulates four causal components, as described in Table 3, and corresponding directly to the prompts provided to students in the “selective pressure” explanation guide.

I examined the two aspects of quality, the articulation of causal components and their warrant, separately. Doing so separates a group’s ability to understand that an appropriate explanation for the problem should be couched within the theory of natural selection (articulation) from their ability to make sense of particular data (warrant). This distinction is important because it is quite possible that students could recognize finch survival as due to differences in individual traits and their possible advantages, but that different groups could determine that different traits were selected for. Explanations claiming that larger beaked birds were more able to survive because of their ability to crack tribulus seeds are the best explanations, from a normative view and in terms of the available data in the investigation environment. Still, any explanation framed in terms of selection of a plausible trait that varies among individuals indicates an understanding of the core ideas of the theory, and would be an improvement in students’ reasoning in comparison to previous studies (e.g., Bishop & Anderson, 1990; Greene, 1990; Settlage, 1994).

Explanations were first scored for articulation, receiving one point for each of the four causal elements they articulated (Table 3). This articulation score could range from 0, if no causal claims were made, to 4, if a complete natural selection explanation was articulated. Explanations received an articulation score for each

TABLE 3
Rubric for Scoring Overall Quality of Explanations

Causal Element	*Articulation (Stated Claim)*	*Warrant (Data to Support Claim)*
Environmental pressure	Describe change in some factor of the environment that could cause a pressure (e.g., drought).	Data showing change in factor over time (e.g., rainfall).
Individual effect	Explain how environmental change affects finches (e.g., less seeds available).	Data justifying link between environmental change and claimed effect (e.g., seed charts).
Differential trait	Identify a trait that distinguishes survivors and casualties (e.g., big beak).	Data that compares this trait in live versus dead finches during affected time period (e.g., beak size graphs during ‘77).
Selective advantage	Claim a causal mechanism for the advantage to individuals that possess the trait (e.g., big beaks can crack tribulus seeds).	Provide behavioral data that links the trait (e.g., big beak) to a function that individuals with that trait can perform that others cannot (e.g., crack seeds).

element if they stated any cause for that element, it did not have to be accurate. The warrant of each claim was assessed by examining the data that each group looked at during their investigation. Table 3 describes the criteria for determining warrant for each causal element. Notice in particular that claims of selective advantage were considered warranted only if data that directly demonstrated the mechanism were available. In the finch problem, such data existed only for the currently accepted scientific explanation that birds with larger beaks could crack open tribulus seeds. Therefore, the only way to score the maximum possible points on both dimensions was to articulate this normative explanation.

This notion of warrant comes from Toulmin's (1958) argument structure. *Warrants* are the reasons given to justify particular claims. In Toulmin's scheme, warrants can come from a variety of sources, including previously justified claims. Here, I consider the data that students looked at to be the source of judgments of warrant, which is more like Toulmin's idea of "backing." My use of warrant here is related to Kuhn et al.'s (1992) notion of inferential validity, except that in their study, inferences about the meaning of data were valid only if they were correct from a normative perspective. I use it here to mean that an inference is plausible and reasonable from the data that students looked at. This is a less strict standard than Kuhn et al. applied, but this seems most appropriate to determine if students are applying the theory of natural selection in a reasonable way to explain this problem. An important aspect of Kuhn et al.'s scheme that I borrowed here is that inferences cannot be warranted by only a single datum. In this problem, students had access to a large number of inscriptions of aggregate data, such as tables or graphs. A single such aggregate graph could be used to warrant a claim. On the other hand, the data about finch behaviors was organized in the investigation environment as field notes about individual birds. A single note was deemed insufficient warrant for a claim, although two or more field notes were considered sufficient warrant. Example explanations in the results section should clarify how warrant was assessed.

All of the explanations for survival written by a group were coded for articulation and warrant. I selected the explanations to be coded as described previously, to score only explanations about survival. Twenty-three explanations from the 19 groups were thus scored independently and blindly by myself and a research assistant unfamiliar with the project. Interrater agreement on articulation and warrant scores was 80%, with disagreements resolved through discussion. The highest scoring explanation from each group, combining articulation and warrant scores, was used in all subsequent analyses.

Causal Coherence

Causal coherence is an important epistemic feature of explanations, and specifically scientific explanations. Here it is a measure of groups' ability to meet the epistemic goals of their inquiry. Causal coherence may also be independent of specific conceptual knowledge. That is, even if groups are unable to articulate natural selection explanations, they may yet strive to produce causally coherent explanations for their inquiry questions. This requires a measure less stringent than warrant, one that assesses the logical coherence of students' causal explanations independent of their biological accuracy.

I derived a measure of coherence from a technique developed to assess the causal cohesion of stories (Trabasso, Secco, & Van Den Broek, 1984), how well stories "hang together." Following Trabasso et al. students' explanations to the finch problem were decomposed into propositions of causes and their effects. From these propositions a network was constructed of the causal relations students' claimed in their explanations. This network was then analyzed to determine the central causal chain of the explanation. Coherence was measured as the ratio of propositions in the central causal chain to the total number of propositions in the network, resulting in a score from 0 to 1. It is important to keep in mind that this is a relative measure of coherence, and it is independent of the content of claims. It measures the extent to which students are explicit about how claims relate to each other.

Figure 3 shows an example of the technique. An explanation is first represented as a numbered list of causal antecedents and consequences, on the left in Figure 3. The graphical network on the right in the figure is constructed by adding each new proposition to the network unless its equivalent has already been added (e.g., line 15 in Figure 3 refers to previously stated things). Solid lines represent explicit causal connections in groups' text, and dashed lines represent implicit connections. Notice that this scheme can represent branches in students' explanations. In this example, the students claim that physiological differences (line 5) and physical size (line 9) could each provide an advantage (line 8). These are separate claims about traits, but they logically cohere in the explanation. Note also that they happen not to be supported by data in the finch investigation environment. So, in the overall quality score students would get credit for articulating a selective advantage, but not for making a warranted claim.

Use of Data

A second epistemic criterion I was interested in here was groups' citation of data to support specific claims. Students cited evidence for their explanations by copying data from the finch investigation environment, and then pasting that data as evidence for a selected component of an explanation in ExplanationConstructor. There are two aspects of students' data use I was interested in. First, did students explicitly cite data at all? Were they sensitive to the epistemic criterion of evidentiary support for causal claims? The data that students cited within their best explanation, as determined by the overall

1. The factor in the environment exerting a pressure is... Lack of rain.
2. It creates a loss of food for the finches.
3. A comparison of young finches in 73 to 77 shows a decrease in births,
4. a complete lack of young surviving. (see graph)
5. Phisiological differences such as metabolism
6. would allow certain finches to attain different food sources
7. or more of a food source,
8. and thus give them an advantage.
9. The physical size of a finch also affects this
10. because a smaller finch eats less food to sustain itself.
11. Finches which do not even eat plants could still be affected
12. because their prey would decrease.
13. But, their decrease would be much lower.
14. Also, stored up energy, fat,
15. the amount of which determined by the above factors,
16. would enable a finch to survive the duration of the lack of food.

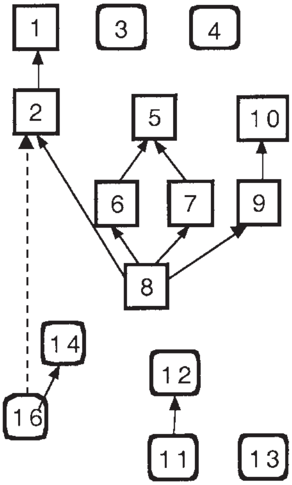


FIGURE 3 An example of the causal coherence coding scheme used here, from group RCS. The numbered lines on the left are the verbatim text of the explanation, as numbered propositions. On the right is the resulting causal network as articulated in the propositions. The square nodes are on the central causal chain (really a network). The coherence score is the ratio of nodes on the central chain (8) to the total nodes in the graph (15), or 0.53.

quality score, were compared to the causal claims in each element. If students cited any data at all that could be deemed relevant to their claim, then they were scored as having cited evidence to support that claim. The determination of relevance was made in terms of students' claims rather than a normative scientific view. Groups did not have to cite sufficient data to support a claim, or even what a biologist might consider to be the most appropriate data. This analysis of relevant data citations is not intended to quantify how well groups used data in their explanations, but to assess students' sensitivity to the need to use evidence to support their claims. Citations of data were judged independently by two raters, with an initial agreement of 89%. All disagreements were resolved through discussion.

A second issue concerns whether students cited sufficient and appropriate data to support a claim. This is a complex determination for students to make, and professional scientists struggle with this issue constantly. Still, students' citations of data for particular causal claims reflects their understanding of the patterns in that data. For example, to make an argument for a particular trait being selected, such as beak length, requires showing differences in that trait across time. Therefore, citing one comparison of beak differences in only a single time point would be relevant, but insufficient, support for a beak difference claim. Judgments of sufficiency thus reflect both conceptual and epistemic aspects of explanation. The judgment of warrant reflects whether students *looked at* enough data to warrant a specific claim, but does not measure whether students *cite* that data as evidence. The citation of data for key claims was quite low in these explanations, as I will show in the following, so I did not pursue an analysis of the sufficiency of cited evidence.

Results

The quantitative scores of overall quality and coherence are intended to portray the patterns of performance of groups across all three classes. I organize these results in terms of the three specific questions guiding my analyses, and I use several examples of explanations to illustrate the nature of groups' performances here in ways that the coding scores cannot easily convey. One note on presentation: Many groups recapitulated earlier explanations for the finches death in their explanations for finch survival, whereas others did not. The overall quality assessments of these latter groups included these explanations for finch deaths, although those parts of the explanation are not always visible in the following examples.

Groups' Articulation of Natural Selection Explanations

I expected that the domain-specific aspect of explanation guide prompts would help groups articulate explanations within a natural selection framework. The guides should focus them on articulating a claim about the trait being selected for, and prior work on people's apparent concern with plausible causal mechanisms (Brem & Rips, 2000; Koslowski, 1996) suggested that most groups would make some claim for selective advantage. At the same time, while guide prompts should help students to decide what kind of data would be important to explain, the prompts themselves were not specific enough to help students resolve ambiguity in the data. So, warrant scores should be lower than articulation scores. In particular, warrant is most likely to be lowest for the claims of selective advantage: only one claim, larger beaks, could possibly be supported by available data. As both Koslowski and Brem and Rips pointed out, however, the absence of data is usually not enough to shake students' preference for plausible causal mechanisms.

Table 4 shows the number of groups earning articulation and warrant scores for each causal component of the natural selection framework. Fourteen of the 19 groups, 74%, articulated an explanation that included all four of the causal elements of the scoring rubric (Table 3). This suggests that the prompts in the explanation guides focused students on the important causal elements to explain, particularly in-

TABLE 4
Number of Groups Articulating, Warranting, and Citing Data for Claims

(N = 19)	*Environmental Pressure*	*Individual Effect*	*Differential Trait*	*Selective Advantage*
Articulation	18	18	15	14
Warrant	18	18	11	6
Data	14	15	5	5

dividual differences in traits, and the selective consequences of those differences. More than this, groups were able to interpret those prompts sensibly, by generating biologically plausible claims for selected traits and their advantages, as shown in Table 5. Also, 12 of the groups, 67%, selected the "selective pressure" guide for their survival explanations, suggesting that they understood the theory of natural selection as being relevant for explaining why the birds survived.

What stands out in Table 4 is the abrupt drop in warranted claims made for both differential traits and their advantages. In comparison to the articulation scores, this suggests that groups understood that they should be using the theory of natural selection to explain survival, but fewer were able to interpret the complex data in the finch environment. The example in Figure 3 shows one group's, RCS (students used their first initials to name their groups), efforts to try to make sense of the ambiguous patterns they observed in the environment. They have articulated all of the elements of a natural selection explanation. In fact, they propose several potential traits being selected for, which suggests some confusion about the various data they have looked at, including scatterplots of individual variation in weight, beak size, and other traits, and field notes showing what the finches ate during different seasons, before and after the drought. RCS illustrate the difficulty in solving this kind of problem, and the tentative voice of their explanation, that metabolism or other factors "would" enable certain finches to survive and others not, indicates their uncertainty.

In contrast, other groups were more sure of their results. (Italicized phrases in these examples represent guide prompts included for context. These are included only when groups used prompts as sentence stems for a particular explanation component, which did not often happen. All spelling and punctuation are otherwise verbatim):

> *The factor in the environment exerting a pressure is …* the drought. The drought exerts pressure because it makes less water available for the plants and finches. The plants can't reproduce because of the shortage of water. Thus producing less food for the finches who in turn begin to compete for food. The trait selected by the pressure of the drought is weight. This is because with less food, the finches began to lose weight. The heavier finches before the drought had an advantage over the lighter ones when the drought

TABLE 5
Claimed Differential Traits and Their Advantages

No. Groups[a]	*Trait*	*Claimed Advantage*
5[b]	Longer/bigger beaks	Able to crack open seeds
3	Weight	Able to live off of fat
2	Foraging skills/knowledge	More able to find food
1	Beaks and wingspans	Stronger and faster
1[b]	Beak and weight (strength)	Strong enough to crack seeds
1	Small, agile	Steal food from larger birds
1	Metabolism	Attain more [energy] from food
1	Small beaks	Able to remove seeds from between spikes
1	Beak length (unspecified)	
1	Changed diet	
2	No trait mentioned	

[a] Two groups offered multiple traits in one explanation, one group gave none. [b] Warranted trait and advantage claims.

> occured. If the heavier finches were either fatter or more muscular they could survive better. The fatter ones could live off their fat, thus needing less food while the muscular finches could obtain the food better due to their physical superiority. (BKJ, “selective pressure,” Quality [summed articulation and warrant scores] = 7, Coherence = 1).

In most respects, BKJ gave an excellent explanation. It was very clearly articulated and coherently laid out the argument for why weight was selected for. It was also consistent with available data. The average weight of the surviving birds went up, because birds with bigger beaks tend to be generally bigger. From a normative perspective, BKJ honed in on the wrong trait. On the other hand, there was no data in the finch environment to contradict their weight claim. They were simply unable to produce data to justify their claimed advantage for weight.

MDB also articulated a complete natural selection explanation, when their separate explanation for the finch deaths is included, although not as clearly as BKJ did:

> The smaller, and more agile finches are able to take the food, which only a magni finch can open. The larger, and more sluggish finches cannot move fast enough to elude the magni finches. Both the large, and small ground finches are not able to open the food, which is left. Some finches are smaller and more agile others are larger and more sluggish physical differences allow the smaller faster ones rob the magni finch. (MDB, “niche divergence,” Quality = 6, Coherence = 0.40)

MDB's choice of trait, small size and agility, was based on a single field note in the database of notes in the investigation environment that described a smaller finch stealing seeds from a larger finch (this single note represents observed behavior during this event). They have generalized from this single observation an explanation for the population, although other field notes would weaken the claim by more directly supporting the claim for larger beak size. Therefore, MDB's claim for the trait and its advantage were judged unwarranted. BKJ's claim for weight, on the other hand, was based on aggregate population trends observed over time and was thus judged warranted.

Six groups articulated complete and fully warranted explanations for the finches survival (Table 4). They were the groups who saw the available field notes on finch eating behaviors and connected those birds who were cracking open tribulus seeds to individual birds on scatterplots of the population. Compared to alternative explanations, such as those offered by BKJ and MDB, they highlight the key difficulty in this problem: coordinating the field notes of finch behaviors with the other available representations of aggregate data showing individual variation along a particular trait, or trait distributions in the population during a specific season. I return to this in the discussion.

These data show that, overall, groups were successful in articulating explanations in terms of the theory of natural selection, but less successful in interpreting the specific data of this problem. They suggest that guidance about the theoretical framework of natural selection can help students attend to relevant aspects of the explanation, such as individual traits, but that additional guidance about the kinds of traits that might matter for a particular organism in a particular environment is needed to help students make better sense of particular data. I stress, though, that the finch investigation was these students' first opportunity to apply the theory to an investigation.

There were differences in overall quality between classes as summarized in Table 6. Independent *t* tests showed that one of the regular level classes had significantly higher warrant scores than the honors class. No other significant differences between classes were found. It is possible that the honors students were more accustomed to situations in which it was clear what the right answer was and how to get it, and there is some evidence that students focused on getting the right answer often fare poorly in inquiry settings (e.g., Tobin, Tippins, & Hook, 1995). I am hesitant to make too much of the difference, however, with such a small sample. No group in any class scored less than 2 on either the articulation or warrant scores, reflecting the fact that all of the groups successfully explained the cause of the finches' deaths as the drought causing a catastrophic loss of food through lack of seeds.

Epistemic Aspects of Groups' Explanations

The second question I posed in this study was whether groups could meet the two focal epistemic criteria for causal coherence and evidentiary support. Before describing these results, let me clarify how these criteria were framed for students. Recall that in discussions immediately prior to the start of the finch investigation students in each class argued that explanations had to cohere, and that claims require evidence. Also, at the start of the finch explanation, the teacher emphasized that groups were to explain how the birds were able to survive—that their explanations were to be causal. He also made the use of data an explicit requirement, although he did not explain what it meant to "use data" to support an explanation. There is some evidence from our field notes that he encouraged students to explain their interpretations of data to him. Because students' data usage is simpler, I start with that.

Students looked at a lot of data in this problem, including charts, graphs, and text field notes. Over the course of their investigations, groups examined an average of 37 (*SD* = 9.74; min = 20, max = 61) unique pieces of data. Many of these data were looked at several times over the course of an investigation, as students repeatedly tried to make sense of data and reassess earlier data in light of new ideas about the problem. Over three fourths of the groups cited data to support their claims of environmental pressure and its effects (last row, Table 4). Groups were clearly able to discover the drought during 1977 and the resulting lack of seeds for the finches to eat, so it may not be surprising that most groups cited these data. The data for these elements were clear and compelling, and students probably felt sure of their importance.

Although Table 5 shows a range of claimed differential traits, only 5 of the 15 groups who claimed a differential trait cited the data on which they apparently based their claim (Table 4). A similar pattern is seen in the citation of data for claims of advantage. Whether a claim was warranted did not seem to affect data citation. Only two of the six groups who made a warranted claim for the advantage of larger beaks cited the crucial field note data to support the claim, even though they had examined that data. Of the other eight groups who claimed an advantage for some other trait, three groups cited supporting evidence. It may be that students did not see the need to cite data for these claims, or perhaps they were unsure of what data to cite. I consider these possibilities further in the discussion, as they

TABLE 6
Overall Quality Scores by Class

	Honors (n = 7)		*Regular 1 (n = 5)*		*Regular 2 (n = 7)*		*Overall (N = 19)*	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Articulation	3.14	1.07	3.60	0.55	3.57	0.79	3.42	0.84
Warrant	2.14	0.38	3.60*	0.55	2.86	0.90	2.79	0.85

*$t(10) = 5.49$, $p < .01$.

highlight the tangled relation between epistemic and conceptual aspects of constructing causal explanations for complex problems.

Causal Coherence

The other major epistemic aspect of these explanations is their internal causal coherence. Did groups articulate coherent causal explanations for this problem? Figure 4 shows the distribution of coherence scores for all groups. The mean coherence score was 0.70 (*SD* = 0.28). Explanations spanned the full range of coherence values, from zero to one. The single explanation scored as zero did not make any causal claims whatsoever. There were five groups whose explanations were scored as fully coherent, a score of one. There was wide variation in the number of propositions in students' explanations (5–20), but longer explanations were not more or less coherent than shorter ones, or more likely to be warranted. The mean coherence score and the distribution of coherence scores across groups shows that most groups wrote coherent causal explanations. Their explanations attempted to live up to their criterion, stated before the finch inquiry, that explanations had to make sense.

Besides providing an overall sense of students' ability to articulate causal explanations for this problem, using ExplanationConstructor, the coherence scores distinguish more and less coherent explanations, and highlight what detracts from causal coherence in explanations. There were two common flaws that lowered students' coherence: the lack of clear causal language, and unconnected causal claims.

Clear causal language. One clear, but not surprising, difference between the high and low coherence explanations is the use of language that explicitly marks

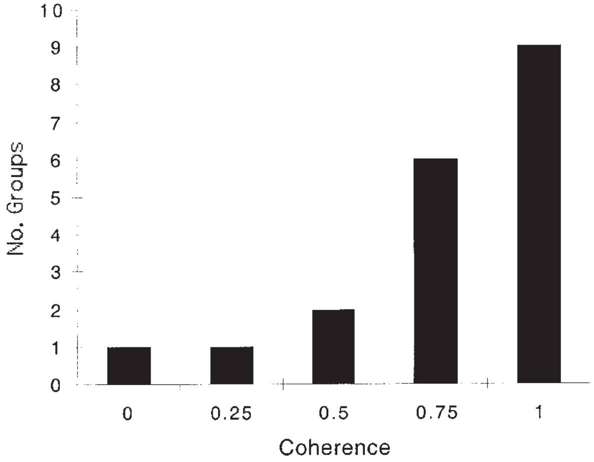


FIGURE 4 Distribution of coherence scores for all groups.

causal claims: “because,” “caused,” “thus,” “due to,” and so on. These words make clear what students are claiming and explicitly connect claims together. Consider again the example of BKJ, shown in Figure 5 as it was coded for coherence: Plants cannot reproduce “because” of the drought, “thus” producing less food which “in turn” leads to competition, and so on. They are very clearly connecting individual claims together. Another example is shown in Figure 6, from the group KAM. KAM are perhaps not as eloquent as BKJ, but they are quite clear in their causal claims. They also do not include any extraneous information.

There is a marked lack of such language in the low coherence explanations. Without these causal connectives, it is difficult to infer students’ causal meaning. Lower coherence explanations typically include a sequence of statements, some of which are causal claims, but with no effort to connect the statements together as a coherent whole. MDB, quoted earlier, are one example. They intimate their causal thinking without being clear, especially about how large size leads to sluggishness. Only a single group had a coherence score of zero, MJM, because they failed to make any causal claims whatsoever:

> *The resources over which individuals compete are* … seeds from local plants, most of which other finches have broken open or are in the process of eating. For evidence, we found in the field notes, that during dry ’77 gf32 was foraging and snatched away a seed from another finch. Some of these resources are found in different locations. Like cactuses are found in different places than other plants. (MJM, “character divergence,” Quality = 4, Coherence = 0).

MJM’s account is merely a sequence of observations, without any explanation of what the observations imply about why some finches died and some did not. MJM were the only group to not make any causal claims whatsoever in an explanation. The rest of the groups at least attempted to explain their observations.

Unconnected claims. Given the definition of coherence I applied here, claims that did not connect to the central causal chain of an explanation lowered its coherence score. It is not necessarily the case that such claims were irrelevant, but that groups failed to integrate them with the rest of their explanation. RCS, for example, proposed several different, unrelated causal mechanisms for why some finches survived, including metabolism, size, type of food eaten, and stored fat (Figure 3). Their inability to connect these claims, to spell out their consequences for birds’ survival, detracted from the coherence of their explanation. It is certainly the case that a lack of clearly causal language contributes to unconnected claims, but that does not fully account for their presence in this sample. Unconnected claims arise in several groups’ explanations as two or more partial explanations that are not integrated into a coherent whole, as with RCS.

1. **The factor in the environment exerting a pressure is...** The drought.
2. The drought exerts pressure
3. because it makes less water available
4. for the plants and finches.
5. The plants can't reproduce
6. because of the shortage of water.
7. Thus producing less food for the finches
8. who in turn begin to compete for food.
9. The trait selected by the pressure of the drought is weight.
10. This is because with less food,
11. the finches began to lose weight.
12. The heavier finches before the drought had an advantage over the lighter ones when the drought occured.
13. If the heavier finches were either fatter or more muscular
14. they could survive better.
15. The fatter ones could live off their fat,
16. thus needing less food
17. while the muscular finches could obtain the food better
18. due to their physical superiority.

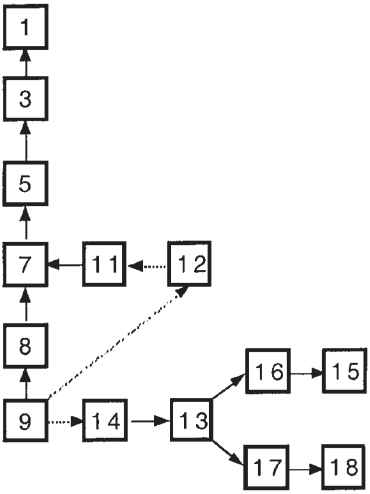


FIGURE 5 BKJ's explanation and coherence network. Overall quality score = 7. Coherence = 1.

Relations Between Conceptual Understanding and Coherence

The final question in my analyses concerns how students' epistemic understanding of the demands for explanation interacts with their conceptual understanding of the theory of natural selection as applied to the finch problem. The coherence score provides a measure of groups' ability to satisfy the epistemic criterion for causal coherence, apart from whether or not they wrote a fully warranted explanation. One of the difficulties here is that it is hard to know if groups' lack of coherence stems from their not being able to figure out why finches were able to survive or from an incomplete understanding of the epistemic demands of the task. If an understanding of the data available is required to write a coherent causal explanation of what happened to the finches, then there should be a clear correlation between coherence and warrant scores. Highly coherent explanations should have high warrant scores, and explanations with low coherence should also have low warrant scores.

1. The rainfall was much lighter, 0-20 inches.
2. the shortage of rain
3. caused the finch's staple plant to deplete.
4. the finches with bigger beaks survived
5. because they could easier,
6. because their bigger beaks help them to adapt to the new diet.
7. it looks like the ones with bigger beaks survived the shortage of portulaca,
8. because their big beaks helped them to crack open shells of different plants.
9. The ones with shorter beaks could not addapt.

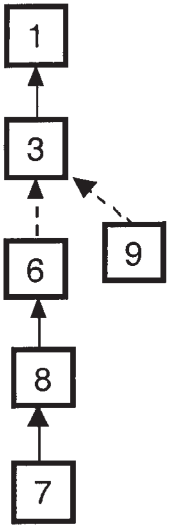


FIGURE 6 KAM's explanation and coherence network. Overall quality = 8. Coherence = 1.

Figure 7 shows the relation between causal coherence and overall warrant. As warranted claims increase, the causal coherence of explanations also increases, Spearman's $\rho = .47, p < .05$. Explanations with lower warrant can also be coherent, but Figure 7 shows that coherent explanations are likely to be more warranted ones. Simply articulating all of the components of a natural selection explanation, whether warranted or not, was not enough to make an explanation coherent, as shown in Figure 8. This makes sense, because an understanding of the epistemic demand for causal coherence is, in itself, unlikely to enable such an explanation to be generated unless some reasonable causal mechanism can be found. Note from Figure 7, however, that the tipping point, as it were, for coherence appeared to be the articulation of a warranted claim for the trait being selected. Groups that made warranted inferences about the differential trait were just as likely to articulate coherent explanations as groups that also made warranted claims about the selective advantage of their claimed trait.

DISCUSSION

I have gone to some length here to disentangle students' ideas about natural selection and how it applies to a specific problem from their ideas about what

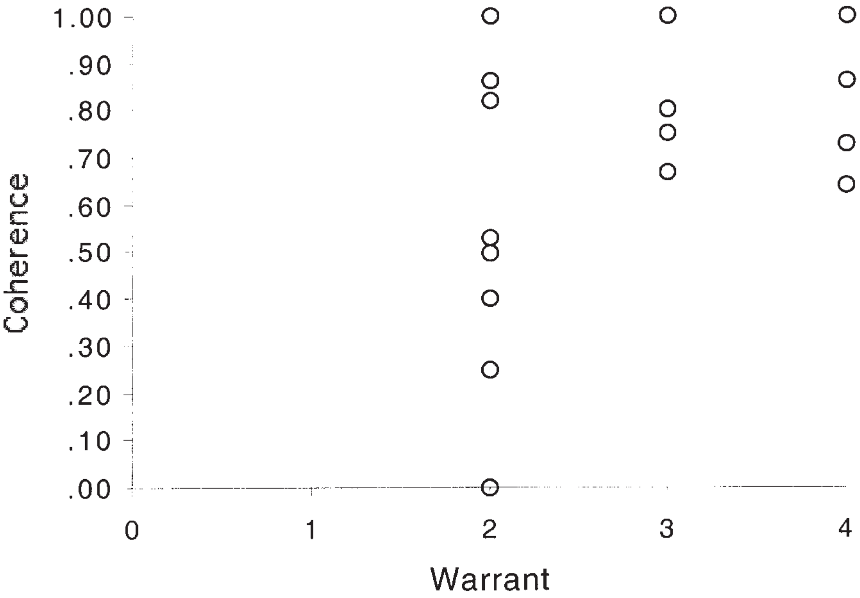


FIGURE 7 Coherence as a function of the overall warrant of explanation (overlap from the same scores on each dimension are not visible on the graph, *N* = 19).

counts as a good scientific explanation, as these are evident in their written explanations for this finch problem. I interpret these findings in three broad ways, that I will take up here in turn. First, explicit guidance about the form of the products of inquiry seems to help students to construct useful products from their inquiry. Second, there is analytic value to trying to explicitly disentangle conceptual and epistemic understanding and how these two aspects of inquiry relate to each other in specific problems. Third, the ability of software programs to provide more explicit epistemic guidance may be limited, and it may be more fruitful to focus on classroom discourse practices and how student-generated artifacts could support them.

Groups' Success at Explanations

The overall pattern of quality and coherence in these explanations for the finch problem shows that groups were successful in articulating explanations for why

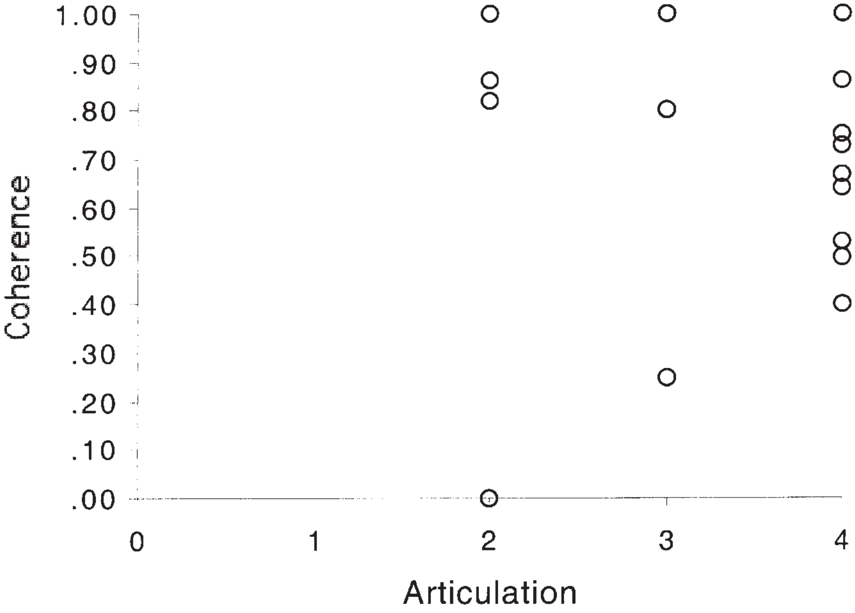


FIGURE 8 Coherence as a function of articulation of natural selection explanation components (overlap from the same scores on each dimension are not visible on the graph, $N = 19$).

some finches could survive a catastrophic drought in terms of the theory of natural selection. Also, these explanations were largely causally coherent, reflecting students' understanding of a central epistemic criterion for scientific explanations. Given that students typically have a hard time explaining problems of natural selection, ExplanationConstructor's guides appear to have helped groups in these classes explain what happened to the finches.

Previous research on students' conceptions of natural selection has consistently found that students overlook or do not understand the role of individual variation in natural selection (Bishop & Anderson, 1990; Greene, 1990; Lawson & Thompson, 1988; Settlage, 1994). Students tend to view individuals in populations "typologically," as equally representative of the traits of the population (Greene, 1990). Yet, over 80% (15 of 18 in Table 4) of the groups here claimed that some difference between individuals, some variation, caused certain finches to survive when others died. Students also went beyond identifying differential traits by proposing causal mechanisms for the selective advantage of those traits. Even when such claims were not warranted from the data, they were biologically plausible causes (Table 5). Thus, students here not only attended to individual variations, they worked to explain why particular differences mattered for survival. They applied the theory of natural selection to the problem.

There are three aspects of the scaffolding provided here that aided groups' articulation of natural selection explanations for the finch problem. First, the problem was framed, both within the software and by the teacher, as an effort to explain why some birds died and others survived. This framing immediately highlighted individual variation by partitioning individuals into one of two groups, fatalities or survivors. Second, access to finch data within the investigation environment was structured through interface tools that highlighted important domain strategies, such as comparisons of physical traits across groups (e.g., live vs. dead) or across time periods (Reiser et al., 2001; Tabak et al., 1996). Thus, the problem framing and students' access to data encouraged students to focus on variation between sub-groups of individuals.

This focus was reinforced by ExplanationConstructor's explanation guides. The general principle behind explanation guide design is to highlight the causal components of important domain theories. In this case, guide prompts were specifically framed to highlight the connection between individual differences and environmental factors, to highlight the notion of differential. The explanation guides made it clear that individual differences, variations, had consequences for the finches' ability to survive, a trend reinforced in the data. Using the selective pressure guide as an example again, the guide prompted students to explain the trait selected by the pressure and to articulate its advantage (Table 2). Groups here were largely able to propose a plausible trait that differentiated the survivors from the fatalities. They were explicit about trait differences that determined survival, which is at least a precursor to overcoming typological thinking if not a definite sign of it.

Other efforts to use prompts to support students' inquiry, and explanation in particular, have also been somewhat successful. Coleman (1998) found that students' use of general, metacognitive prompts during collaborative attempts to explain photosynthesis led to better explanations. Coleman also found, however, that students often did not know when to refer to the prompts they were given, or necessarily how to respond to them. Davis and Linn (2000) found that general prompts that encouraged students to reflect on their progress during an investigation fostered students' integration of normative conceptions of heat and temperature with their prior personal knowledge. Explanation guide prompts may encourage a similar level of reflection. Explanation guide prompts outlined specific disciplinary explanatory frameworks that could be applied to the problem, but at a level of generality that demanded students consider how they might be fulfilled. They provided clues about what needed to be explained while remaining open-ended.

It is important to understand that there was nothing in the content of the explanation guide prompts that could tell students what specific data were important, or how to interpret a particular graph or other data. Rather, prompts suggested the kinds of data that would be important to construct an explanation (e.g., data that will show a change in the environment that exerts pressure on the finch population). Also, prompts here guided students' interpretations of the kinds of relations to look for among the data. For example, a key relation was the connection between aggregate representations of the physical characteristics of finches (e.g., the beak sizes of live and dead birds during the dry season of 1977), the physical traits of individual finches, and the field note data about individual finches' behaviors. Computer traces of groups' data explorations show that they attended to the "position" of individual birds in aggregate comparisons of physical traits of survivors and casualties. The explanation guide prompts made such an interpretive move especially salient. Students had difficulty connecting such physical trends with the behavioral data in part because of the difficulty in comparing multiple sources and forms of data, and the ambiguity inherent in that data. Explanation guides, as implemented here, do not support such problems of interpretation. These guides can, however, focus students' explanations, and efforts to seek data for explanations, in conceptually relevant ways.

Disentangling Epistemic and Conceptual Aspects of Inquiry

The notion of an epistemic game (Collins & Ferguson, 1993) provides a good metaphor for thinking about how students' scientific epistemologies manifest themselves during inquiry. Playing a game is active. These students' explanations shed light on their "active" scientific epistemologies, in contrast to professed epistemological ideas assessed through most of the research in this area. Previous research on students' ideas about the nature of science has almost exclusively focused on surveys or interviews about their ideas on philosophical or social aspects of science (see Lederman, 1992, for a review), or abstract conceptual ideas about the nature of theories and experiments (Carey et al., 1989; Carey & Smith, 1993; Smith, Maclin, Houghton, & Hennessey, 2000). Only rarely have students' scientific epistemologies been assessed through concrete explanatory tasks (e.g., Driver et al., 1996). It is also likely that students' explanations for the finch problem reflect more personal epistemic views of themselves as students and learners of school science (cf. Hammer, 1994; Hogan, 2000). Such personal epistemologies are not obviously reflected in students' explanations; nonetheless, these explanations for the finch problem provide an important lens on students' epistemic understanding of the actual practice, the epistemic game, of science and scientific explanation.

The point I want to make is that although students' conceptual understanding of the data they explored and of the theory of natural selection clearly affected their efforts to explain this problem, that alone does not explain what groups did here. Rather, groups' explanations reflect two epistemic aspects of their work. One is the effort to generate plausible causal mechanisms to explain ambiguous data. A second, less positive, aspect evident here is that students seemed to view data as something to be explained, but not necessarily as a necessary component of an argument. Examining how these played out here will lead to pedagogical implications for promoting effective science inquiry.

Plausible Causes for Ambiguous Data

Students here were actively trying to make sense of complex data, to generate causal accounts that could order the data. Students proposed causal mechanisms that could explain trends in observed data (e.g., differences in weight, or beak length), or generated plausible causal mechanisms in the absence of data (e.g., possible advantages to greater weight). What groups' finch explanations show is not only that they were more or less able to articulate causal accounts, but that they recognized that as an epistemic goal of the inquiry. First, nearly all groups articulated a complete, if not accurate, explanation. Also, except in rare cases, lower coherence scores in this sample were not due to a lack of causal claims but to the fact that partial explanations were often disconnected from each other within the entire explanation. Students' performance here is consistent with recent work showing that adolescents and adults prefer plausible causal explanations even in the face of scarce evidence (Brem & Rips, 2000) or contrary evidence that does not suggest a plausible mechanism (Koslowski, 1996).

Deanna Kuhn (1993; Kuhn et al., 1988) has argued that most people hold an epistemic stance that does not distinguish between the theories that they believe and the evidence that may support or refute those theories. An alternative view proposes that children and adults can, in fact, distinguish causal claims from evidence in various settings (Brem & Rips, 2000; Koslowski, 1996; Sodian, Zaitchik, & Carey, 1991; Tschirgi, 1980). The data here generally support this latter view, and suggest that students hold an epistemic view of explanation as an effort to articulate causal mechanisms to explain data. More than half of the groups here articulated warranted claims about possible differentiating traits (Table 4). It is important to note, though, that even the groups whose claims for traits were not warranted from the data they examined were still making inferences from that data (as suggested in Table 5). The claim for small size and agility, for example, was based on data from a field note in the database. The claim for metabolism was generated as an attempt to explain observed differences in weight, and so on. These were causal claims about, and distinct from, observed data.

Thus, the difficulty for students here does not seem to be an inability to distinguish claims from data, but that the data themselves were complex and hard to explain. Moreover, it seems unlikely that students held strong beliefs about particular causal mechanisms, such as metabolism, prior to the problem, and then ignored data counter to those beliefs (cf. Kuhn et al., 1988). Students were free to generate a new explanation whenever they wanted to, either before or after looking at some, or even all, of the data. Students may have had prior ideas about the importance of certain physical characteristics, such as weight being associated with strength (see Table 5), that were triggered as possible causal mechanisms in response to the data they saw. They were doing the opposite of seeking confirming data for prior beliefs. They were generating possible causal mechanisms for the available data.

Students did not go as far as one might wish. They did not, for example, hold the lack of confirming data for claims of advantage to be, effectively, counter evidence. This could be because students did not see a lack of data as problematic, or took aggregate data showing trait differences as *de facto* evidence for the trait's advantage. The former view is more purely epistemic than the latter, in that the decision of whether one has enough evidence for a claim or not is partially contingent on an understanding of what might count as appropriate evidence (Zeidler, 1997). Yet, even in such cases students were doing what they had been asked to do: to use the theory of natural selection to explain the problem. They used the theoretical suggestion that trait differences have selective advantages to advance causal claims. The value of this is that students' explanations thus provide a starting point for examining conceptions about the theory as it applies to this problem. Moreover, the effort to advance plausible causal mechanisms for incomplete data is legitimate scientific practice (we might not have Darwin's theory without such an effort), and a necessary first step in the classroom to engage students in the evaluation of claims in relation to data and to overarching domain theory.

The Role of Evidence in Explanations

Groups' explanations suggest that they were actively reasoning about data when constructing their explanations. Yet, students were often not arguing with data. Remember, the lack of data citation is not due to students' failure to consider the data while constructing their explanations. If that were so, students would not have made as many warranted claims as they did. The pattern of data citation, common for claims of environmental pressure and individual effects, less common for differential traits, and rare for selective advantage, suggests interesting possibilities about students' ideas of the role of data in supporting an argument.

One possibility is that students just did not see explicit evidence as a necessary part of their explanations. Instead, they may have viewed data simply as the means for generating claims, not substantiating them. Such an orientation to explanation fits the goals of typical school science. Getting the story right is what counts, not necessarily backing up that story with evidence. This orientation could have operated here despite our efforts to the contrary. On the other hand, if students here had a general notion that evidence was not important, data citation should have been uniformly low across all components of their explanations. Also, class discussions immediately prior to the finch investigation indicated that students were aware of the need for data to support claims.

Another possibility is that students cited only the data they felt certain that they understood. Groups were much more likely to cite the rainfall data that established the drought (i.e., environmental pressure) and the seed data that supported the claims of individual effects. These are the data that establish the reason for why so many finches died. These are also the two components of the explanation that all of the groups agreed on. Every group concluded that the drought caused the lack of seeds that led to mass starvation. As groups tried to understand which of the several possible traits might have been selected for survival, and why, the claims they made diverged and citations of data decreased.

Another way that uncertainty of the meaning of data might inhibit explicit citation is that students may have simply run out of time. Many groups constructed explanations in a piecemeal way, filling in parts of an explanation when they felt they had something important to record, as in the vignette of Evan, Franny, and Janie. Groups often, however, waited until a particular explanation was finished before retrospectively citing data. In such a strategy, figuring out the story for yourself takes precedence over documenting the evidence that leads to the story. There is some evidence that groups ran out of time here, although careful observations were taken of a limited number of groups, only one in each class, making it hard to claim lack of time as a definitive reason.

My analyses do not permit any of these possibilities to be ruled out. Rather, they suggest that each of them are likely. Each may explain the performance of some of the groups in these classes. The explanations themselves document the struggles that students had to understand these data, and field notes and video records corroborate this. At the same time, even those groups who arrived at the same explanation as the biologists who watched the events unfold (Grant, 1986) did not cite the data that they apparently understood.

Explanations as a Measure of Epistemic Practice

Ultimately, what do these analyses uncover about students' epistemic understanding of inquiry and explanation? These explanations illuminate some of the active epistemic practices students engage in during complex inquiry. Students here made an effort to articulate causal mechanisms to explain data, and they were sensitive to the criterion that such causal accounts should cohere. Students' ability to articulate a coherent explanation relied in part on their ability to make sense of the available data, of course, although groups who could propose a reasonable causal mechanism for survival were able to coherently articulate it even when it was not entirely warranted. They also have some idea that it is important to show the evidence for a claim. At the same time, epistemic strategies that would have been helpful here, such as viewing a lack of data as a weakness for a claim, and a reason to seek an alternative account, were rarely evident.

More importantly, the effort to disentangle conceptual and epistemic aspects of explanation is an explicit move to recognize that both play integral roles in students' attempts to learn science, and especially to learn science through inquiry. Beyond simply recognizing this, however, these analyses are an attempt to understand students' epistemic practices as they play out in their scientific work. Hogan (2000) argued that students' beliefs about themselves as science learners play a more important role in their science learning than do their ideas of professional science. There is evidence to support this view (Hammer, 1994; Linn & Songer, 1993; Songer & Linn, 1991; Windschitl & Andre, 1998). On the other hand, students' ideas about formal science may manifest themselves quite differently through their own scientific inquiry and sense-making than they do in formal, often abstract surveys and interviews.

Students' explanations are only suggestive of the possible epistemic resources (Hammer & Elby, 2001) students used to construct them. To fully understand students' epistemic beliefs requires more direct probes of them, both in the abstract and in relation to their inquiry activities. Still, as artifacts of students' understanding, these explanations show that students do, in fact, see that explanations should articulate causal accounts, and that causal claims must be based upon available data. It is almost certain to be the case that asking students to interpret their own inquiry performances will illuminate aspects of their scientific epistemologies that surveys or abstract interviews do not reach, while also providing pedagogically rich opportunities for reflection. To that end, these explanations provide important artifacts of student understanding for reflection and critical evaluation.

Pedagogical Implications for Inquiry-Based Science

In light of groups' performance on this finch investigation, I draw two main pedagogical implications from these analyses. One is that there are limits to the kinds of support that technology can provide for students' inquiry, especially in relation to helping students evaluate the fit of their claims to available evidence. The second, related to the first, is that the explanations that students produced here and other kinds of artifacts can provide an important resource for epistemic discourse in the classroom, and such a discourse is probably necessary to develop students' epistemological understanding of science.

Limits to Technological Scaffolds

The version of ExplanationConstructor used in this study seems to have provided some clear support for students' articulation of coherent causal explanations of natural selection, the targeted goal of the investigation. Students struggled most in coordinating the relevant available data with their causal claims. The ability of ExplanationConstructor or other software programs to help in this effort is probably limited. A flaw in this version of ExplanationConstructor is that it was not immediately visible when looking at an explanation whether data had been cited for each component. Students may have been unaware in some cases whether they had cited data for claims. On the other hand, the pattern of data citation in these explanations was related to apparent difficulties in data interpretation, rather than being directly hindered by a slightly cumbersome interface. The lack of immediate visibility of the relation between data and claims may partially account for the low amount of data citation, but not all of it. These findings led, however, to revisions to ExplanationConstructor to make the relations between evidence and explanations immediately visible, and more salient.

The component structure of the explanation guides in this version of ExplanationConstructor may have hindered some groups' efforts to write coherent explanations. Although overall, groups were largely coherent, the groups with lower coherence seemed to rely on prompts for connecting language even though the prompts themselves did not really provide that. Subsequent revisions to ExplanationConstructor moved explanation guides out of the space where students write explanations, to encourage students to write more clearly narrative accounts. The results from these revisions will be the focus of subsequent analyses.

More importantly, the software cannot indicate to students whether their claims make sense, are consistent with available data, or have been sufficiently supported with relevant evidence. Given that one of the main goals of inquiry-driven science reforms is to develop students' understanding of the nature of science, computer coaching on the correctness or coherence of their explanations may not be desirable. It seems unwise to replace the teacher-as-authority with the computer-as-authority when the goal is for students to develop for themselves criteria and standards for scientific knowledge claims. In professional science, evaluations of arguments and theories are largely social; other people have to be persuaded by one's claims and evidence (Latour & Woolgar, 1986). Such processes of persuasion should become more prominent in the classroom as well. Although this software enabled students to construct rich artifacts of their understanding of this problem, students' use of such technologies is insufficient to transform the prevailing science discourse of most classrooms.

Epistemic Discourse Around Artifacts

The act of investigation and explanation itself is not always enough to help students see that a particular claim is not warranted by the data they observed, or that they have failed to connect two partial explanations, or that they have not given evidence to support a claim. Pushing on these aspects of performance seems to require a public classroom discourse focused on students' explanations, in both conceptual and epistemic terms. Conceptually, such a discourse would focus on whether the claims of selected traits are sensible given what is known about the theory of natural selection, and about the organisms being studied, in this case small birds. Through a public conversation about the relative merits of groups' various explanations, the teacher, for example, could bring out how unlikely it is that birds as small as these finches could live off of their body fat for 6 months. Those groups that saw and successfully interpreted the beak size data in relation to the field notes of birds' eating tribulus seeds would make their argument. Through such a reflective discourse (vanZee & Minstrell, 1997), students' understanding of the finch problem would be clarified and could be connected to the theory of natural selection. This is, in fact, usually what happens in classrooms using this curriculum (Reiser et al., 2001; Tabak & Reiser, 1997).

Epistemically, such a discourse would focus on the coherence of groups' claims, and how any particular claim can be judged as warranted. Interventions with an emphasis on epistemic discourse seem to be successful at developing students' epistemological ideas about science (Rosebery, Warren, & Conant, 1992; Smith et al., 2000). The classrooms in both of these studies, for example, placed a heavy emphasis on holding claims accountable to evidence. Each also seemed to include multiple, ongoing opportunities for epistemically focused discourse about the bases and justifications for specific claims, methods for generating data to answer questions, and evaluating the fit of current ideas to data. A striking feature of these studies is that the interventions they describe each occurred over an extend period. Rosebery et al.'s study spanned an entire school year. Smith et al. described children's epistemological ideas after 6 years of science instruction with the same teacher!

Artifacts like the explanations that students constructed here can support such a discourse. They make students' ideas explicit, and organize their ideas in epistemically relevant ways. Students' journals, for example, link explanations to questions, and evidence to causal claims. Thus, the artifacts can set the parameters of the discussion. Has a specific question been answered? Is the proposed explanation coherent? Is it well-supported? Comparisons of different groups' explanations can occur along the same lines. Similar to my study, Bell and Linn (2000) found that structuring students' investigations around a debate, where they had to defend one or another hypothesis about light, encouraged students to explain the data that they saw. During debates, students were able to use the common structure of their artifacts to organize their arguments. As with the students here, however, the creation of the artifacts in themselves were insufficient to push groups to explore the limits of their arguments, as arguments or with respect to their understanding of light. An outstanding issue then, is how to make the best use of the artifacts that students generate from inquiry to build on their understanding about both particular science concepts and about science as a practice.

I should point out here that the teacher and I organized a post-investigation discussion in which groups critiqued each other following the finch problem, but it was largely unsuccessful. Students were not well prepared to critique each others' work, and there was a definite sense that they, and possibly the teacher too, did not see the value in explaining and evaluating each others' work. Students tended to congratulate each other for drawing the same conclusions, or they talked about other things. This emphasizes, especially in light of previous work, the importance of a well-structured epistemic discourse guided by teachers. Such a discourse relies on changing the norms operating in classrooms regarding the nature of scientific discourse. Lemke (1990) described how the typical science discourse in classrooms is authoritarian, and evidence is often framed as being objective rather than theory-laden. These attributes of typical discourse probably play a large role in developing students' scientific epistemologies by the time they reach high school, and are generally opposed to the epistemology underlying inquiry. This history, together with the data here and from other studies, suggests that the development of epistemic discourse in science classes takes time. Even so, groups' performances on this finch problem suggest productive epistemic starting points for such a discourse.

CONCLUSIONS

The practices of constructing and defending explanations are now seen as central to scientific practice, and as necessary means to developing students' conceptual and epistemic understanding of science (Driver, Newton, & Osborne, 2000; Duschl, 1990; Kuhn, 1993). I began this article by suggesting that focusing students on the nature of the products of inquiry, namely causal explanations, could guide their inquiry processes. The evidence here suggests that students adopted explanatory goals, primary among them the need to explain data. Moreover, their efforts to make sense of data were grounded within these explanatory goals. Thus, epistemic scaffolds grounded within domain-specific guidance appear effective in focusing students on important aspects of the products of their inquiry, and encourage an orientation to data as something to be explained. This orientation to data as something to be explained is a valuable outlook for students to take, and is a necessary step toward developing students' epistemologies of the nature of scientific inquiry and explanation. These specific inquiry experiences are not enough, however. Instead, the artifacts that students generate from their inquiry should be used to focus classroom discourse on the relations between causal ideas and data that relate to them. Such a discourse would not only develop students' understanding of key conceptual ideas, but enables the development of epistemological understanding grounded within real efforts to make sense of phenomena, to make scientific knowledge.

As researchers, analyzing the artifacts from students' inquiry can provide important clues to students' epistemological ideas as they play out during their inquiry. Such knowledge is crucial for the development of effective inquiry-oriented pedagogy. In the analyses presented here, I have taken a step toward developing a more grounded account of students' epistemic practices during inquiry. These analyses have shown that there are specifically epistemic aspects evident in students' performances that researchers and educators should attend to, while making clear that epistemic and conceptual understanding are tightly interrelated. Thus, students' explicit use of evidence to support claims is both constrained by their ability to make sense of it and guided by their ideas about what evidence is necessary to support a claim. Further research needs to relate such artifact analyses to students' discourse during inquiry, broader classroom discourse norms, and other assessments of students' epistemological ideas about science. Such research will provide a needed foundation for inquiry-based reform efforts that not only help students to learn science concepts, but about scientific practice too.

ACKNOWLEDGMENTS

This research was supported by a grant from the James S. McDonnell Foundation to Brian J. Reiser and James P. Spillane.

I am very grateful to Brian J. Reiser for his guidance and support of this work. Many people have contributed to BGuILE and influenced this work, including Iris Tabak, Brian Smith, Franci Steinmuller, Angie Agganis, Eric Baumgartner, Tammy Porter, Eric Fusilero, Richard Leider, T. J. Leone, and Renee Judd. I especially thank Iris for her collaboration on this study. I am grateful to the teacher and students who so graciously participated in this research. Thanks also to Charles Framularo and Kate Muir, at UCLA, for their help with reliability analyses.

This research was conducted as part of my doctoral dissertation at Northwestern University.

REFERENCES

AAAS. (1992). *Science for all Americans*. Washington, DC: Author.

Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education, 22*, 665–701.

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.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*, 415–427.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573–604.

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education, 68*, 493–503.

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.

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

Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences, 7*, 387–427.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

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

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287–312.

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

Duschl, R. A. (1990). *Restructuring science education: The importance of theories and their development*. New York: Teachers College 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: Lawrence Erlbaum Associates, Inc.

Grant, P. R. (1986). *Ecology and evolution of Darwin's finches*. Princeton, NJ: Princeton University Press.

Greene, E. D. (1990). The logic of university students' misunderstanding of natural selection. *Journal of Research in Science Teaching, 27*, 875–885.

Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction, 12*, 151–183.

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88–140.

Hogan, K. (1999). Relating students' personal frameworks for science learning to their cognition in collaborative contexts. *Science Education, 83*, 1–32.

Hogan, K. (2000). Exploring a process view of students' knowledge about the nature of science. *Science Education, 84*, 51–70.

Jackson, S. L., Stratford, S. J., Krajcik, J., & Soloway, E. (1994). Making dynamic modeling accessible to precollege science students. *Interactive Learning Environments, 4*, 233–257.

Klahr, D., Dunbar, K., & Fay, A. L. (1990). Designing good experiments to test bad hypotheses. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 355–401). San Mateo, CA: Morgan Kaufman.

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

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

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

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*, 285–327.

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

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

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching, 25*, 733–746.

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

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex.

Linn, M. C., & Songer, N. B. (1993). How do students make sense of science? *Merrill-Palmer Quarterly, 39*(1), 47–73.

Mayr, E. (1988). *Toward a new philosophy of biology*. Cambridge, MA: Harvard University Press.

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

Ohlsson, S. (1992). The cognitive skill of theory articulation: A neglected aspect of science education? *Science & Education, 1*, 181–192.

Reif, F., & Larkin, J. H. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. *Journal of Research in Science Teaching, 28*, 733–760.

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 classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–305). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

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

Roth, W. -M., & Roychoudhury, A. (1994). Physics students' epistemologies and views about knowing and learning. *Journal of Research in Science Teaching, 31*(1), 5–30.

Sandoval, W. A. (1998). ExplanationConstructor [Computer software]. Evanston, IL: Northwestern University.

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. *Journal of the Learning Sciences, 4*, 131–166.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences, 1*, 201–238.

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.

Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching, 31*, 449–457.

Shute, V. J., Glaser, R., & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 279–326). New York: Freeman.

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.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–766.

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

Suthers, D., Weiner, A., Connelly, J., & Paolucci, M. (1995, August). *Belvedere: Engaging students in critical discussion of science and public policy issues.* Paper presented at the 7th World Conference on Artificial Intelligence in Education (AI–ED95), Washington, DC.

Tabak, I., & Reiser, B. J. (1997). Complementary roles of software-based scaffolding and teacher–student interactions in inquiry learning. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Computer Supported Collaborative Learning '97* (pp. 289–298). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tabak, I., Smith, B. K., Sandoval, W. A., Agganis, A., & Reiser, B. J. (1996, April 8–12). BGuILE: Supporting inquiry in a learning environment for biology. Paper presented at the Annual Meeting of the AERA, New York.

Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific support for biological inquiry. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *Proceedings of Intelligent Tutoring Systems '96* (pp. 288–296). Montreal, Canda: Springer-Verlag.

Tabak, I. E. (1999). *Unraveling the development of scientific literacy: Domain-specific inquiry support in a system of cognitive and social interactions.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Tobin, K., Tippins, D. J., & Hook, K. S. (1995). Students' beliefs about epistemology, science, and classroom learning: A question of fit. In S. M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 85–110). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Trabasso, T., Secco, T., & Van Den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1–10.

vanZee, E. H., & Minstrell, J. (1997). Reflective discourse: Developing shared understandings in a physics classroom. *International Journal of Science Education, 19*, 209–228.

Weiner, J. (1994). *The beak of the finch: A story of evolution in our time*. New York: Knopf.

Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching, 35*, 145–160.

Wood-Robinson, C. (1995). Children's biological ideas: Knowledge about ecology, inheritance, and evolution. In S. M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 111–130). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. *Science Education, 81*, 483–496.