Students' uses of data as evidence in scientific explanations

William A. Sandoval
Graduate School of Education & Information Studies
University of California, Los Angeles
sandoval@ucla.edu

Paper presented at
the Annual Meeting of the American Educational Research Association
Seattle, WA, April 19-23, 2001.

Acknowledgements

This work was supported in part by the James S. McDonnell Foundation and by a grant from the UCLA Academic Senate's Council on Research. The findings and opinions expressed here do not necessarily represent the views of these agencies. Support was also provided by UCLA's undergraduate Student Research Program, and I am grateful to Jason Finley for his work on these analyses. I am indebted to the BGuILE team, headed by Brian Reiser and Jim Spillane, and to the teachers and students who took part in this study.

Sandoval, W. A. (2001). Students' uses of data as evidence in scientific explanations., Paper presented at the Annual Meeting of the American Educational Research Assn. Seattle, WA, April 10-14.

Abstract

A central goal of scientific explanation is to account for patterns of data. An important way to assess students' abilities to construct scientific explanations is to examine how they use data as evidence. A good deal of cognitive research has explored how students respond to specific pieces of data and evaluate them in terms of a current theory or belief. This work finds that, from a normative view, students often ignore data that they ought to consider when evaluating claims, or assimilate such data in ways that do not damage their current theories. Researchers differ on whether or not such reasoning reflects an epistemological stance that is fundamentally non-scientific, or is essentially reasonable and consistent with scientific practice. This study examines high school students' efforts to use complex, multi-faceted data sets to construct causal explanations of natural selection phenomena, consistent with the theory of natural selection. The study seeks answers to two questions. What data do students select to use as evidence for their explanations? How do they refer to specific data to justify particular causal claims? Content analyses of explanations examined how students referred to data as evidence: the features of inscriptions (e.g., graphs, field notes) they referred to, and the justifications they gave for the importance of specific data. Students largely cited relevant, but insufficient, data for claims, and preferred numerical data over textual field notes. These and other uses of evidence indicate not just students' understanding of what specific data mean, but their ideas about what counts as persuasive evidence. These findings highlight the need for instruction focused on epistemic criteria for good explanations, in addition to the conceptual relations between data and potential causal mechanisms.

Sandoval, W. A. (2001). Students' uses of data as evidence in scientific explanations., Paper presented at the Annual Meeting of the American Educational Research Assn. Seattle, WA, April 10-14.

Introduction

A central goal of scientific explanation is to account for patterns of data. An important way to assess students' abilities to construct scientific explanations is to examine how they use data as evidence. A good deal of cognitive research has explored how students respond to specific pieces of data and evaluate them in terms of a current theory or belief (Chinn & Brewer, 1998; Klahr, Dunbar, & Fay, 1990; Koslowski, 1996; Kuhn, Amsel, & O'Loughlin, 1988). This work finds that, from a normative view, students often ignore data that they ought to consider when evaluating claims, or assimilate such data in ways that do not damage their current theories. Researchers differ on whether or not such reasoning reflects an epistemological stance that is fundamentally non-scientific (Kuhn et al., 1988), or is essentially reasonable and consistent with scientific practice (Chinn & Brewer, 1993; Koslowski, 1996). This study examines high school students' efforts to use complex, multi-faceted data sets to construct causal explanations of natural selection phenomena, using the theory of natural selection. The study seeks to answer two questions. What data do students select to use as evidence for their explanations? How do they refer to specific data to justify particular causal claims? Both questions are answered through content analyses of students explanations, including an examination of the sufficiency of data that students cite to support their claims and the justifications they give for the importance of specific data.

The view of scientific practice I rely on here is more sociocultural than cognitive. I borrow from theories of situated learning (Lave & Wenger, 1991) and cognitive apprenticeship (Collins, Brown, & Newman, 1989) the idea that learning science is an apprenticeship into ways of thinking. Such ways of thinking manifest themselves in the kinds of practices that students can engage in while they attempt to conduct authentic scientific inquiry. My interest here is therefore not to measure how much knowledge students might acquire through their inquiry, but to document how students construct meaning from data and how they view the relationship between data and possible interpretations of its meaning. The data that students select to include in explanations are likely to reflect their ideas about what data are important in understanding a particular natural phenomenon or event, what data are important to report in an explanation, and what data they feel like they understand. There is a complex entanglement here of students' epistemic understanding of scientific explanation and argumentation and their conceptual understanding of particular domains, theories, and problem situations. One effort of this study is therefore to determine if students' patterns of data citations in explanations to particular problems can help us to disentangle the influence of epistemic and conceptual understanding in students' efforts to scientifically explain natural phenomena.

The second goal of this study is to determine the extent to which students' use of data in their explanations can be considered scientific. The cognitive science literature strongly suggests that it is not, although this view has been challenged in recent years (Brem & Rips, ; Chinn & Brewer, 1993; Koslowski, 1996). These challenges suggest that all of the flaws that students manifest in reasoning about data, including seeking only confirming data, ignoring contradictory data, conflating correlation with causation, and more, can be seen in historical cases of professional science and can be interpreted as rational inquiry. The view of science as a practice, rather than simple knowledge acquisition, has more recently led to research suggesting that learning science entails learning to interpret discipline-based inscriptions that scientists use to represent natural phenomena and claims about them (e.g., Roth & McGinn, 1998). Much of this work has focused on students' interpretation of specific inscriptions, such as graphs, as indicated by how students explain what an inscription "says." Yet, scientists do much more than explain particular inscriptions. The coordination of evidence with theory that has been the focus of much of cognitive research, is recast in this sociocultural view as the coordination of inscriptions for the rhetorical purpose of making a persuasive argument. Clearly, this requires that students understand how particular theories can be used to interpret evidence. Yet, it also requires an understanding of the rhetorical purpose of scientific explanations and the criteria by which such explanations might be judged. That is, successfully coordinating multiple sources of data in a scientific explanation requires more than a conceptual understanding of specific theories and concepts, it demands an epistemic understanding of explanations.

How much of such an understanding do students have? Students' epistemologies of science have been widely studied for several decades, and the general impression is that students have simplistic ideas about what science is (see the review by Lederman, 1992). Importantly, students do not seem to see science as a process of explanation, and often seem to think that data are answers in themselves (Carey & Smith, 1993; Driver, Leach, Millar, & Scott, 1996). The great bulk of studies of students' scientific epistemologies have, however, been isolated from contexts of students' actual science practice. This study is an effort to understand how students' epistemic understanding manifests itself in their practice. To provide a framework for analyzing students' use of inscriptions in their own explanations for natural events, I borrow from Latour's (1987; Latour & Woolgar, 1986) accounts of professional laboratory work as an effort to coordinate inscriptions into persuasive arguments for the construction of scientific "facts." To the extent that students argue through the inscriptions they select as important evidence, Latour's work suggests that their explanation can be seen as scientific. The criterion here is not some normative notion of what science *ought* to be, but *actual* scientific practice.

Science as literary inscription

Latour notes a paradox in professional scientific work: scientists' efforts center around the production of argument texts that entail essentially literary processes of inscription designed to persuade readers of "facts," but such arguments are persuasive only to the extent that all signs of persuasion are lost or forgotten in the final argument. That is, successful persuasion occurs when inscriptions can be seen as essentially congruent to the "facts" that they construct. Much of this effort at literary inscription for scientists is an effort, then, to generate inscriptions that can be seen as having some direct correspondence to the facts of interest. For example, in his study of a laboratory in the Salk Institute, Latour describes how the scientists in the laboratory talked about the difference between peaks on a graph as if they were the chemical substance (TRF) they were trying to find, rather than lines on paper produced by a machine designed to generate such peaks (Latour & Woolgar, 1986). During the effort to construct a fact like TRF such an inscription is quite tenuous as the peaks on the graph may or may not signify the particular object being sought. Yet, by the end of the process, the inscription is taken by participants as directly representing TRF.

In school science such ideas of tenuousness are rarely raised. Instead, teachers typically communicate an authoritarian view of science facts (Lemke, 1990), and the use and meaning of inscriptions seems to be a rare issue. Textbooks include many and various inscriptions, but these are rarely explained or referred to in ways that are common to scientific practice (Roth, 19??).

Inquiry-based instructional approaches to science learning often involve students in the creation of inscriptions to communicate results from experiments. This is hard for students (Krajcik, Blumenfeld, Marx, Bass, & Fredricks, 1998), although even middle school students seem aware of the rhetorical power of effective representations (diSessa, Hammer, Sherin, & Kolpakowski, 1991). In the present study, students were not asked to construct inscriptions themselves, but to make sense of inscriptions of various kinds provided to them as a result of data queries. For example, while investigating why some finches on a Galapagos island survived a catastrophic drought and others did not, students could compare the beak length of live and dead birds during a particular season. Such a query would generate a scatterplot for each group, showing the beak length of each individual live or dead finch. Students task was thus to understand the patterns in such inscriptions and others (including histograms, bar charts, pie charts, tables, and textual field notes), and use them to construct an explanation for a particular problem.

Given Latour's characterization of scientists' view of inscriptions as directly showing scientific "facts" and the backdrop of typical science instruction just described, students' own use of data in their explanations may similarly take for granted what particular inscriptions demonstrate. That is, we might expect that students let the data "do the talking," without explicitly guiding their audience to a particular interpretation of specific inscriptions or combinations of inscriptions. To the extent that students explicitly refer to features of inscriptions for rhetorical purposes, to make their claims more persuasive, they can be said to be engaging in scientific argumentation. Flaws in students arguments, such as citing irrelevant, contradictory, or insufficient data for a particular claim may not result from an essentially unscientific way of thinking (e.g., Kuhn, 1993), but instead indicate students' lack of conceptual understanding within particular domains (e.g., Bowen, Roth, & McGinn, 1999) and their inexperience with scientific inscription practices. An important way in which this study is different from earlier studies of students' use of scientific inscriptions, is that the explanatory context here is explicitly rhetorical. This focus enables an extension to analyses of students' ideas about what particular inscriptions "say" to what those inscriptions mean with respect to a particular argument.

Method

Setting

This study took place in four introductory high school biology classes, taught by two teachers, in a suburb of a major Midwestern city. At the time of the study, the school was approximately 85% Euro-American students, with less than two percent on a school lunch program. Mr. Goodson[1] taught two honors level classes of introductory biology, with a total of 52 students. Students were admitted in the honors class on the basis of a written test unrelated to biology. Mr. Burns taught two regular track introductory biology classes with a total of 35 students. A total of 87 students participated in the study (43 girls; 44 boys). Mr. Goodson was working with our research team for the third consecutive year. Mr. Burns was working with us for the first time.

Over four weeks, each class worked concurrently through the same evolution unit, developed as part of the Biology Guided Inquiry Learning Environments (BGuILE) project. Reiser, et al., provide a comprehensive description of the project's approach, curricula developed, and several classroom studies. For this paper, I only briefly describe the nature of student work as it relates to the data sources analyzed here. The evolution unit was centered around two, week-long, computer-supported inquiry projects. The first investigation was to explain how some finches on an island in the Galapagos archipelago were able to survive a catastrophic drought during 1977. The second investigation, conducted a week after the finch investigation, asked students to explain how the bacteria that causes tuberculosis, *M. tuberculosis*, develops resistance to antibiotics (Reiser et al.,).

During both of these investigations, students collaborated in groups of three or four to explore large sets of data in the computer investigation environments and use them to explain each problem. Groups recorded their explanations in a software tool we developed, ExplanationConstructor, that guided their explanation in several ways (see Sandoval & Reiser, 1998). Of particular importance to the current analysis, ExplanationConstructor provides facilities for students to link data that they consider to be important evidence for their claims into the text of their explanations. These data are generated in specific investigation environments, and copied into electronic journals students keep in ExplanationConstructor.

Data sources

For this study, we collected the explanations that students recorded for each of the Finch and TB investigations. This paper reports on our analyses to date on students' explanations for the Finch problem. We collected a total of 32 explanations written by 30 groups for the Finch problem. For this analysis, we selected the most complete explanation offered by a group for the problem. Because the goal of each investigation was for students to explain each problem in terms of the theory of natural selection, the explanation that came closest to articulating a natural selection explanation, as determined below, was chosen as the most complete and used for all subsequent analyses.

Analytical methods

For the Finch investigation, students were asked to explain an overarching driving question. For the Finch investigation, the driving question was, "Why are some finches able to survive [a catastrophe] when others cannot?" This is a question of natural selection. Briefly, the theory of natural selection states that individual organisms within a population that possess particular traits are more likely to survive in given environmental conditions. In the wild, this typically means that when a population's environment changes in some drastic way, only some individuals, if any, will possess the traits needed to survive the change. Thus, for the Finch problem, we were looking for students to articulate four major elements of an explanation by natural selection: 1) the change in the environment exerting selective pressure on individuals; 2) the effect that pressure had on individual finches or bacteria cells; 3) the trait that differentiated survivors from casualties; and 4) the selective advantage of the trait.[2].

Overall quality

All of the explanations that groups recorded in their journals, as saved in ExplanationConstructor, were initially scanned using the above natural selection rubric. In most cases, groups wrote a single explanation that best fit the rubric, although occasionally groups had two competing explanations. All of these candidate explanations were scored for overall quality using the above rubric as follows (Sandoval, in review). First, an explanation received a point for each causal element in the rubric that it articulated, for a maximum of four points. I refer to this as the *articulation* score. Then each causal element was scored for inferential validity (Kuhn, Schauble, & Garcia-Mila, 1992) by examining trace logs of the data that students examined during their investigations and deciding if students had examined data that warranted the causal claim. I call this the *validity* score. For groups having more than one candidate explanation to the driving question, the explanation with the highest combined articulation and validity score was considered their best explanation and used for subsequent analysis.

Use of data

The primary focus of this study was to understand how students used data in their explanations to support their claims. Such an analysis is, at heart, an effort to understand the quality of student arguments. Analyses of students' scientific arguments are commonly conducted in terms of Toulmin's (1958) formal argument structure. Toulmin's argument structure includes claims, data, warrants, and backing. This structure has been used to analyze students' argumentative discourse (Kelly, Druker, & Chen, 1998), constructed arguments (Bell & Linn, 2000), and software tools designed to support argumentation (Bell & Linn, ; Suthers, Toth, & Weiner, 1997). Toulmin's framework allows the structure of students' arguments to be analyzed, but it does not allow, by itself, an analysis of the sense or correctness of such arguments (see Driver, Newton, & Osborne, 2000). While a structural analysis such as Toulmin's can provide useful insights into students' mode of argumentation, it is also necessary to understand the quality of those arguments, and this requires understanding such arguments in terms of specific subject matter knowledge. Here, then, we analyzed students' use of data as evidence in their explanations along two major dimensions: sufficiency, and rhetorical use.

Sufficiency

Students cited data by linking specific items into their explanation text. Judgments of the citation of relevant and sufficient data were combined as follows. Our first level of analysis for data use was to determine if students cited any data at all to support their causal claims. This analysis was conducted in terms of the natural selection rubric described above. That is, if students claimed that a lack of rain created an environmental pressure for the finches, did they cite data about rainfall? For this analysis, data that was not judged to pertain to one of the four causal elements of a natural selection explanation were not included. To create a persuasive explanation, of course, requires not just relevant data, but sufficient relevant data to warrant a claim. To analyze sufficiency of data, we created a general rubric for each of the four causal elements of a natural selection explanation (Table 1). Note that this judgment of sufficient data underlies the judgment of inferential validity in the overall quality score. Although the amount of data needed to sufficiently support each causal claim varied both between problems and between causal elements within the two problems, we defined a four level scheme to score the sufficiency of relevant cited data (Table 2).

These levels of sufficiency emerged from our analyses as an effort to capture the range of students' citation of evidence. Students sometimes cited no data at all for specific claims, although cited data was nearly always relevant. Some groups cited only a single item of data, for example, when several items were needed to warrant a claim. Other groups cited nearly enough data to be considered sufficient, but not quite if held to a strict "expert" standard. For example, a claimed differential trait had to be supported by evidence that surviving individuals, as a group, had a given trait and casualties, as a group did not (level 2). Strictly speaking, though, to be fully sufficient such a claim should also show that this pattern did not exist prior to the introduction of a selective pressure (level 3).

Table 1: Rubric used for coding use of sufficient evidence.

Causal Element	Sufficient data to support claim
Environmental Pressure	Data showing a change in the environment during the time selection is claimed to occur.
Individual Effect	Data linking the environmental pressure to a claimed effect on individuals, requires multiple individuals suffering the effect.
Differential Trait	Data showing a trait difference between individuals correlated with survival (i.e., individuals that have it survive, those that do not, die).
Selective Advantage	Behavioral data that links the trait to a function that individuals with that trait can perform that others cannot. Data must compare behavior/function between individuals that survive and those that do not.

Table 2: Scoring levels for sufficiency of cited data.

Level	Definition
0	No relevant data cited for a claim
1	Citation of some relevant data
2	Citation of nearly sufficient, relevant data
3	Citation of fully sufficient data

Rhetorical use

Ultimately, with respect to students' understanding of various inscriptions, and how to argue with them, we are most interested in how students refer to data rhetorically, to make their arguments. We identified three categories of rhetorical use prior to our analysis. At the simplest level, students merely *included* data in their explanations without any exposition as to its purpose, the claim(s) it was intended to support, or why it supported them. A slightly more sophisticated use was to *describe* the data within an explanation, but to do so at a level that still did not explain how the data supported particular claims. At the highest level of rhetorical sophistication, students *interpreted* the data that they cited in terms of its support for particular claims.

The text of each explanation was examined to determine how groups referred to each piece of data that had previously been judged as relevant. Each reference to a particular piece of data (e.g., a graph, a text field note, a DNA sequencing results) was scored in one of the three categories of rhetorical use. This allowed us to examine overall trends across groups, variability within groups, and potential connections between rhetorical uses of data and particular causal elements of an explanation. That is, students may have felt that some data, or some claims, required more explanation than others.

Results

I present results from analyses of students' explanations for the first investigation conducted during the evolution unit, natural selection among finches on the Galapagos island of Daphne Major (see Grant, 1986). Analyses of students' explanations for the second problem are forthcoming. I present these results in three sections. First, I present scores for overall quality to provide an indication of students' performance on the problem. Next, I present the categorical analysis of students' citations of data. Third, I examine how students really use data in their explanations by looking at how they refer to it. This section points out what features of inscriptions students refer to to make their arguments.

Overall quality of explanations

As in previous studies (Sandoval, in review; Sandoval & Reiser, 1997), students here were overwhelmingly able to articulate explanations in terms of natural selection (Table 3). For the finch problem, explanations tend to have two parts. The first part, including the articulation of an environmental pressure and its effect on individuals, essentially explains why so many finches die. The second part, including the claim for a differential trait and its selective advantage, explains how some of the finches are able to survive. As with previous studies, the articulation of this second part falls slightly from the first part.

Table 3: Number of groups articulating inferential valid natural selection claims for the finch problem.

(N=30)	*Environmental Pressure*	*Individual Effect*	*Differential Trait*	*Selective Advantage*
Articulation	30	29	27	27
Validity	30	28	22	5

Sufficiency of data

The overall pattern of data citations is shown in Figure 1. As with articulation, students' citation of sufficient data declines in the more complex causal components of their explanations. One thing is clear, however. Students generally seem to understand that they need to cite some data to support their claims, but sufficiency drops dramatically for claims of differential traits and their advantages. The large number of groups with no relevant data for selective advantage claims has two causes. The first stems from our rubric for deciding whether or not data is relevant, and reflects a feature of the Galapagos Finch investigation environment (and the island). We required students to cite behavioral data to get credit for citing relevant data for claims of selective advantage. In the investigation environment, the only available such data describes finches' eating behaviors, and therefore is only relevant to claims of beak differences. This means that those groups that make other claims for differential traits cannot generate relevant data to support their claims. Still, a large number of groups claimed that beak differences were selected for but did not cite any relevant data. Instead, students relied on aggregate differences in beak size for their claims of selective advantage. That is, they appeared to consider the demonstration of a trait difference as synonymous with advantage. This is less than ideal, of course, but we stress that it represents a good approximation of the question they were asked to explain. Finches survived because they had bigger beaks.

It is important to try to understand what is going on here, as evidenced by the claims students make, and do not make, and the data they cite, or do not cite, as evidence. There is an interplay between students' understanding of the kind of explanations they ought to produce, and their understanding of available data and its significance in relation to their problem.

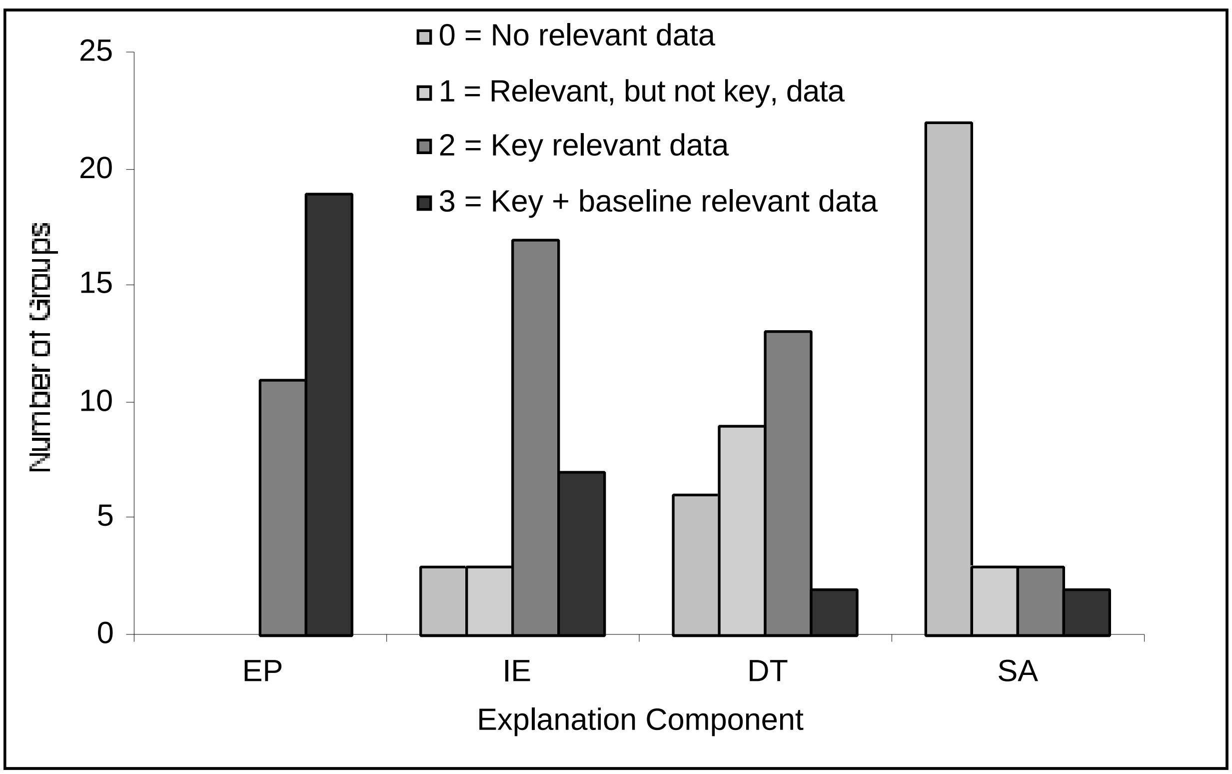

Figure 1: Sufficiency score distributions for each causal component of groups' finch explanation (N = 30). Relevance judged in relation to stated claim. Key data included data relevant to period of selection (e.g., dry season of '76 through dry season of '77). Baseline data included relevant data from period prior to key period.

The overall pattern as shown in Figure 1 indicates that students understand that their explanations require data for evidence. The distribution across causal elements reflects various sources of difficulty within the problem and available data. Also, as noted above, some claims are more easily supported than others. A claim for rainfall as the environmental pressure simply requires a single graph that depicts rainfall over several years. Claims for lack of seeds as the individual effect requires, to be strictly sufficient, that students show that, in fact, the finches eat the seeds that decline during the drought. Most students did not do this, and would probably argue that it is obvious that finches eat seeds. The pattern for data citations about possible differential traits is more spread out because students understanding of the kind of data required to demonstrate such differentiation appeared to vary. About a third of the groups cited nearly sufficient data, documenting a difference in beak sizes between live and dead birds during the drought period. A few groups gained full sufficiency scores by showing that this difference did not exist during the baseline year. Another third of the groups cited at least one graph showing beak size differences, but not during the crucial time period when selection was happening. Eight groups failed to cite any supporting data, but five of these did not produce claims for a differential trait. There seems to be a complex relationship between students citations of data, their understanding of the rhetorical power of (or need for) data in their explanations, and their understanding of what specific data mean with respect to their driving question.

We also computed a sufficiency score for each group by averaging the sufficiency scores for each causal element, as an indicator of each group's citation of data. The mean group sufficiency score was $M = 1.61$, $SD = 0.52$. Overall group sufficiency was strongly correlated with the overall inferential validity of explanations, Spearman's rho = .832, $p < .001$. This is in marked contrast to the previous year's study at this same site (Sandoval & Reiser, 1997), and reflects a much greater emphasis by the teachers in this study on the use of data. Also, not surprisingly, overall sufficiency scores are correlated with higher amounts of data cited, Spearman's rho = .501, $p < .01$. This reflects the fact that in the finch problem sufficient support for claims of differential traits and selective advantage require multiple sources of data.

Rhetorical references to data

We have looked at students' rhetorical references to data at a fine grained level, by scoring each citation of data according to the scheme described above. We have also looked at rhetorical reference by group, to explore relationships between students' general patterns of reference, evidentiary sufficiency, and the inferential validity of their explanations.

Frequencies of rhetorical references

We analyzed students' rhetorical references to data in two ways. Figure 2 shows the frequency of our three *a priori* rhetorical categories across all citations of data for all groups. That is, each data item cited in an explanation was scored. Each citation received a single score, the highest level reference made to it in the explanation. As Figure 2 shows, students were overwhelmingly explicit about their rhetorical use of their evidence. Students were far more likely to explicitly interpret the meaning of data for their claims than they were to simply describe data or merely include it, $X^2(314, 2) = 61.21$, $p<.001$.

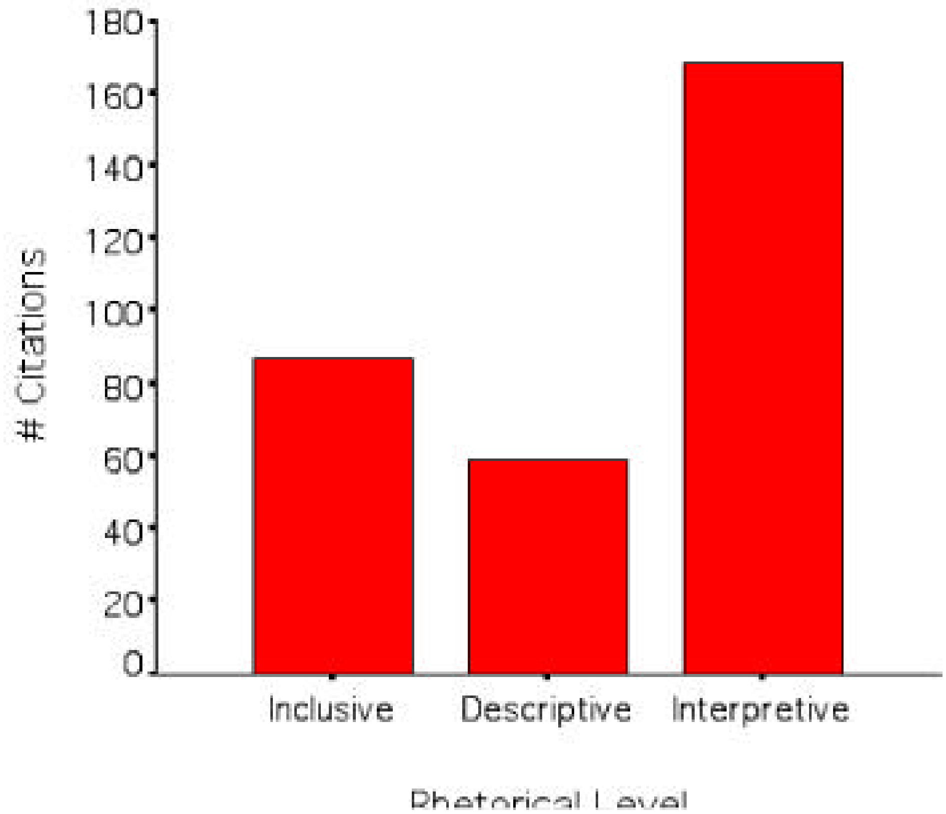

Figure 2: Distribution of rhetorical reference across all citations (N=314).

Simple inclusive references occur when students link data into their explanations but make no explicit reference to it. There is a sense in these cases that students believe that the data speak for themselves, that their import is self-evident. This may be because students actually think this, akin to Latour's descriptions of professional scientists arguing that inscriptions bear a straightforward relation to objects (Latour & Woolgar, 1986). On the other hand, it may occur here because the corpus of data is shared by all of the participants in the classroom, students and teachers. Students may feel that the teacher already knows which data are important (not necessarily the case at all in this instance), and therefore it matters only to cite the relevant data.

Examples of descriptive references are shown in Table 4. These references are typically summaries of what a graph shows, or occasionally an interpretation of the meaning of the graph itself. We consider such interpretations *local* in the sense that they propose a meaning for the inscription, but one that is irrespective of its rhetorical impact.

Table 4: Examples of descriptive rhetorical references in students' explanations.

Explanation text
graph below shows…decrease…in portulaca
during the wet season of 77 there was very little rain. The seasons before and after wet 77 had no rain at all.
shows a slight difference in the average length of live and dead birds
male beak is larger than the females, as shown in the graph
only 20 inches of rain, but normally there is 150-200 inches...
shows that in some dry seasons, the rainfall is very little
the data shows that the finches are often found foraging for these certain seeds
these two charts compare the beak lengths of the live and dead birds…average beak length of the living is around 12.5 inches. The average beak length of the finches that died was about 11 inches.
the finches had previously been seen eating portalaca and cactus seeds (figure 2, 3)

The most sophisticated rhetorical references to data are those in which students interpret the significance of the data for their explanation (see Table 5). These interpretations extend the local meaning of particular inscriptions by attributing persuasive power to them. As such, students' language in these instances is quite different. In our sample, many interpretive references are accompanied by descriptive references to the same data. The following example is typical of this form of reference:

Table 5: Examples of interpretive rhetorical references in students' explanations.

Explanation text
the living birds averaged longer beaks, which supports our theory.
this graph is a little sketchy…males…heavier might be because they were able to gather the most food. On the other hand, they might just be naturally heavier.
as you can see…finches went down and so did the owls. We can tell…from this graph that the owls weren't a factor…
can attribute…finches death to the seed decline because we [k]now that seeds were their main source of food, and their deaths were consistent with the seed decline.
through this chart we see that finch #20, which had a closer to average beak survive, while finch #49, which died had a short beak. This supports our idea…
this is hard to support, since there is no data that shows…
according to our data, the owl population droped at the same time the fnch population dropped, therefore the owls could not be a major influence of the finches drop.
for the living finches the average wing length, leg length and weight was almost exactly the same as the dead finches averages… this shows that the beak length…determined with or not the finch lived.

Such references can be thought of as compound references, in that students rely on the descriptive reference to lend weight to their rhetorical interpretation. We would argue that such compound references are consistent with professional scientific practice. The preliminary coding we present here does not count such compound references explicitly, however. Ongoing analyses will capture this, as such compound references better reflect how students rely on particular inscription features to support their claims (see below).

Patterns of reference

The overall distribution of rhetorical references shown in Figure 2 provides only a coarse view of students' rhetorical use of data. We wanted to understand how possible differences in rhetorical references related the quality of students' explanations, and whether or not rhetorical use was related to sufficiency. As a first step, we computed a rhetorical reference score for each group by simply averaging the individual rhetorical scores for each data citation in their explanations. The mean group rhetorical reference score was $M = 2.17$ ($SD = 0.68$). This mean suggests that groups were generally describing their data, but Figure 3 shows a different pattern. About one quarter of the groups, seven, were mostly including data without explicitly referring to it at all. Another third of the groups (eleven) were mostly or entirely explicitly interpreting the meaning of data with respect to their causal claims. The rest of the groups were a mix. Overall, these data show that most groups have some understanding of the rhetorical power of data in their explanations.

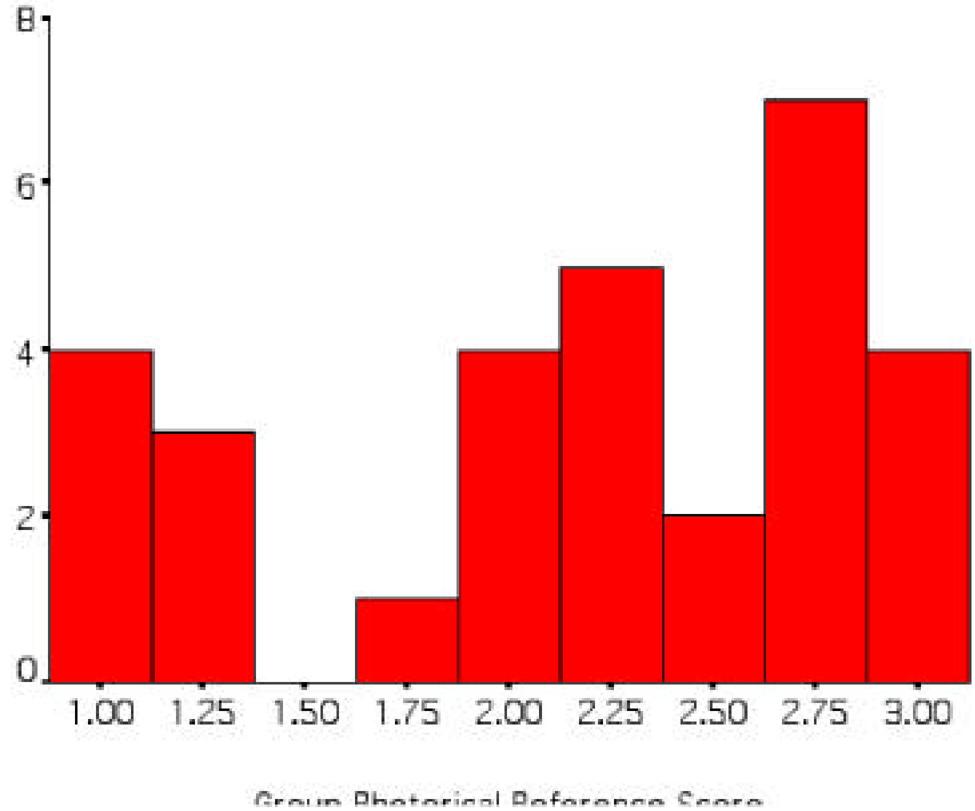

Figure 3: Distribution of rhetorical reference by group (N=30).

We compared these group rhetorical reference scores to overall quality scores and sufficiency scores. We found no strong correlation between groups' rhetorical references to data and the sufficiency of data cited, or to the inferential validity of their explanations. This is not especially surprising. First, the quality of students' explanations was assessed on the basis of their causal claims, and inferential validity was determined by the data students examined, whether they cited that data or not. That is, explanation quality is a measure of how well students were able to figure out what was going on in this problem, and how well they could articulate their conclusions within a natural selection framework. Also, the definition of sufficiency we have used here relies simply on the inclusion of data in an explanation, the lowest level of rhetorical reference.

Interestingly, students rarely referred to a lack of data in order to discount claims. They had been asked specifically by their teachers to consider alternative hypotheses, and they did so. Such alternatives were generally refuted by citing contradictory data (such citations are included in the data about rhetorical reference, but not in the sufficiency scores). On the other hand, many claims for advantage for some trait other than beak length were based on inferences that could not possibly be supported by data because such data was unavailable. One of our hopes is that students would come to see such a lack of data as potential evidence against claims, or at least spur them to consider other hypotheses. As many others have pointed out, however, such a point of view is difficult for both students and scientists to take (e.g., Chinn & Brewer, 1998; Kuhn et al., 1992).

Reference to inscription features

In his analyses of scientific work, Latour describes how scientists refer to specific features of inscriptions to make claims. Educational researchers, particularly Roth (see Roth & McGinn, 1998), have recently considered how students make sense of inscriptions in conversation. Roth and McGinn give an example of how a teacher and his students use a series of pulley diagrams to anchor their talk during a whole class discussion, and how participants in the discussion point to features of the inscription and essentially make claims through them. In our sample of written explanations, such kinds of specific reference to inscription features is less visible, and perhaps more difficult to do in writing than in conversation. The most common feature that students explicitly refer to are numerical quantities (as suggested by Tables 4 and 5). Especially, students are often quite explicit in describing trends in rainfall and seed production, and their correlation. These data are quite compelling within the finch environment and form the bases of students' claims for environmental pressure (i.e., lack of rain) and individual effect (i.e., lack of seeds causing increased competition for food and/or starvation). Although preliminary, our content analyses of groups' explanations suggests that at least some students are also able to point out differences between groups of finches as shown in scatterplots and histograms. When such references are explicit, students report the quantities they are attending to as a way of drawing attention to the features of the inscriptions they find important.

These references to data within students' texts almost exclusively deal with quantitative data. Indeed, students rarely cite available field notes as evidence (a pattern seen in previous studies), even though such data are necessary to establish claims for selective advantage. Many more groups examined such notes, although here only 8 groups cited field note data for their claims of selective advantage. Most groups did not cite any data at all for claims of advantage (Figure 1), although most groups claimed that longer beaks were being selected for, and there were field notes that would support that claim. Instead, most groups seemed to infer mechanisms of advantage to explain the trait they decided was being selected by the drought. Such a strategy is entirely reasonable, of course, especially when data is unavailable to confirm or contradict a claim (Koslowski, 1996). We consider this further in the discussion.

Discussion

In our analyses we have borrowed from Latour's framework of scientific work as literary inscription as a way to understand students' use of data as evidence in explanations for a complex problem. Although our analyses are preliminary, one of the major findings of this study is that students' rhetorical use of evidence in this context has much in common with Latour's accounts of professional scientific practice (Latour, 1987; Latour & Woolgar, 1986). In particular, students here were largely sensitive to the need to cite data to support their claims, and they commonly used their data rhetorically to support their claims. We also claim that a second contribution of this analysis is to suggest a way to extend recent work examining the structure of students' arguments (Bell & Linn, 2000; Kelly et al., 1998) by considering the quality of the data students use to back up their claims.

Showing data vs. using data

There are several aspects of the instructional context created here that support students' use of data. For one thing, the teachers made data citation, or at least inclusion, an explicit performance criterion. This performance demand could be largely responsible for the large increase in data citations in this study compared to a prior one in this same school (Sandoval & Reiser, 1997). Second, software scaffolds available to students during their investigations made the incorporation of specific data into the texts of their explanations easy. Third, students had opportunities during their investigation to critique each others' work, which again highlighted the utility of having evidence to back up claims (for a more complete description of this curriculum, see Reiser et al.,). On the other hand, there was little discussion during this unit about what counted as relevant and sufficient data, nor was there discussion about how to use data in explanations. Thus, these groups' explanations provide an indication of students' understanding of relevant and sufficient evidence within a context in which they are encouraged to use data, but in which it is largely up to them to figure out what that means. For some students here, that means simply including the data they find relevant, and for still more that means summarizing what that data says (Figure 3).

Many students went beyond simply showing data, and explicitly tried to explain why a particular piece of data supported a given claim. Yet, for the more difficult parts of these explanations, the claims for a differential trait and its advantage, students had difficulty citing sufficient data. Not only was it common for students to not cite enough data, some students actually cited data that either contradicted or did not actually support their claim. This is a problem, and it may be due in part to students' commitment to prior beliefs or their refusal to accept contradictory data (cf. Chinn & Brewer, 1998; Klahr et al., 1990; Kuhn et al., 1988). In this case, however, the problem does not seem to be a commitment to prior belief, or an epistemological commitment to confirmatory data. Rather, the finch problem is a really hard problem, and the data students are asked to interpret are complex and ambiguous. This point bears stressing, because the rich complexity of this inquiry context make it quite different from typical cognitive psychology studies in which students evaluate a single piece of evidence at a time, usually with respect to a single causal relation. Here, the data are complex and the explanatory task is complex. As with similar inquiry-based efforts, students lack both the domain knowledge and strategic knowledge to always accurately interpret data (e.g., Krajcik et al., 1998).

All of this is to say that it is important in this context that students be pushed to closely examine the data with respect to their emergent claims. Although we have been successful in creating rich investigative contexts in which students are clearly having some success, it remains difficult to create situations within the broader classroom context in which students challenge each others' claims and evidence, and thus consider in more depth what particular inscriptions really say, and how inscriptions combine to produce converging evidence for or against particular explanations. Based on our analyses of students' use of data here, it seems likely that part of students' preparation for significant investigations in particular domains should include practice generating and "reading" inscriptions relevant to the discipline being studied. This goes beyond simply understanding what, for example, a histogram "says," and extends to knowledge of the situations in which being able to "say" what a histogram says is important (i.e., in order to argue for the distribution of a trait within a population). Such rhetorical knowledge is rooted both in conceptual knowledge of specific domains and practitioners' epistemic understanding of what counts as good evidence in their discipline.

Combining structural and content analyses of student arguments

The analyses here are a first approximation to content analyses that can capture the quality of students' explanations together with their argument structure. Such analyses are crucial to be able to assess whether or not students really get better at constructing good explanations. Toulmin's (1958) argument structure has become popular recently as an analytic tool in science education research, but as Driver and colleagues have recently pointed out such analyses are unable to say whether or not structurally sound arguments actually make any sense (Driver et al., 2000). Bell and Linn (2000), for example, report the frequency with which students provide warrants and backing for their arguments for a preferred hypothesis about light propagation. Their structural analysis provides useful insight into the structure of students' arguments, especially in relation to their innovative inquiry curriculum. For instance, students who provide unique (i.e., invented) backings for arguments were more likely to integrate formal light concepts with their own prior knowledge. This structural analysis cannot say, unfortunately, whether or not such backings are sensible, or reflect an improved conception of light propagation.

Our analyses of the sufficiency of cited data are one way to infuse such structural analyses with a measure of the quality of the content of the argument. Sufficiency explicitly assesses whether cited data (i.e., backing) provides appropriate support, or refutation, of particular causal claims because it is grounded within specific disciplinary theoretical frameworks (evolutionary biology in our case). The sufficiency measure as we have defined and used it here allows for an assessment both of the amount and kind of data students are using as evidence. It can thus indicate the depth or limits to student understanding of important concepts. For instance, the sufficiency measure as defined and used here distinguishes between students that cite data for a differential trait with data only for one group (e.g., live or dead finches; level 1) from students who recognize that a comparison of live vs. dead is the key (level 2, at least).

The analysis of sufficiency presented here remains coarse, however. In particular, it does not distinguish between data that are merely included in an explanation and those that are used rhetorically to support claims. Such a coarseness will suffice only if we presume that the mere inclusion of data reflects an understanding of that data. This may not always be a safe presumption. We separated judgments of sufficiency from judgments of rhetorical reference because we did not want to assume that students do not understand data simply because they do not explicitly refer to it in their explanation texts. That is problematic, although it seems preferable to us to have some indicator of students' use of evidence that is tied directly to the content of their explanations (i.e.., arguments) rather than just their structure.

Conclusions

We have used a perspective of scientific activity as literary inscription (Latour, 1987; Latour & Woolgar, 1986) as an analytic device to understand how students use rich, multi-faceted sets of data as evidence for their explanations of an event of natural selection. This approach has provided a framework in which to examine the quality of students' use of data, in terms of sufficiency and rhetorical effect, with respect to the content of their explanations, rather than merely their structure. In contrast to studies of students' interpretation of inscriptions in conversation, we have analyzed students' rhetorical use of evidence in written explanations. Within our inquiry-based curriculum on evolution in which students were strongly encouraged to include data as evidence for causal claims, students did so. Students quite commonly made explicit rhetorical use of their cited data to make their claims more persuasive, and explicitly referred to aspects of inscriptions that we would claim are scientific, in that they are similar to the practice of scientists. Students pointed out trends in data, they related disparate data sources (e.g., rainfall and seed production), and they pointed out salient comparisons. Our content analyses of students' explanations, although somewhat preliminary, suggest that students may have a bias toward quantitative forms of evidence, although this needs further research.

Our findings here suggest that, much like their intuitive conceptions of particular domains, students' intuitive ideas about explanation and the use of evidence can be usefully built upon. It does not seem in this case that students' citation of inappropriate evidence, or failure to cite evidence, stem from confirmation biases, but rather come from legitimate difficulties making sense of complex data within a new domain. The teacher-led demand here for the citation of evidence encouraged students to cite data for their claims and, generally, to try to use that data persuasively. More attention in the classroom to the development and use of evaluation criteria may help students to better evaluate their developing explanations in ways that support effective monitoring of their conceptual understanding. Such a focus on epistemic cognition (Kitchener, 1983) seems necessary for inquiry-based approaches to science learning to fulfill their promise.

---

[1] All participant names reported here are pseudonyms.

[2] For the finches, the current explanation is that birds with larger than average beaks were able to crack open the hard-shelled seeds that survived an extended drought (Grant, 1986).

References

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

Bowen, G. M., Roth, W.-M., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching, 36*(9), 1020-1043.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*(4), 573-604.

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

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

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35*(6), 623-654.

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: Erlbaum.

diSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: meta-representational expertise in children. *Journal of Mathematical Behavior, 10*, 117-160.

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

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

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

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: combining performance assessments with argumentation analysis. *International Journal of Science Education, 20*(7), 849-871.

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition. *Human Development, 26*(4), 222-232.

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

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

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., & Fredricks, J. (1998). Inquiry in project-based science classrooms: initial attempts by middle school students. *Journal of the Learning Sciences, 7*(3&4), 313-350.

Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly, 39*(1), 74-103.

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

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University 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: Cambridge University Press.

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*(4), 331-359.

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

Reiser, B. J., Tabak, I., et al. (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.

Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: Toward a theory of representing as social practice. *Review of Educational Research, 68*(1), 35-59.

Sandoval, W. A. (in review). Conceptual and epistemological aspects of students' understanding of scientific explanations.

Sandoval, W. A., & Reiser, B. J. (1997). Evolving explanations in high school biology, *Paper presented at the Annual Meeting of the American Educational Research Assn.* Chicago, IL, March 24-28.

Sandoval, W. A., & Reiser, B. J. (1998). Iterative design of a technology-supported biological inquiry curriculum, *Paper presented at the Annual Meeting of the American Educational Research Assn.* San Diego, CA, April 13-17.

Suthers, D., Toth, E., & Weiner, A. (1997). An integrated approach to implementing collaborative inquiry in the classroom. In R. Hall & N. Miyake & N. Enyedy (Eds.), *Proceedings of CSCL'97: The Second Intl. Conference on Computer Support for Collaborative Learning*.

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