COGNITION AND INSTRUCTION, *30*(3), 207–245, 2012

ISSN: 0737-0008 print / 1532-690X online
DOI: 10.1080/07370008.2012.689383

Routledge
Taylor & Francis Group

A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning

Michael J. Ford
University of Pittsburgh

This article identifies aspects of argumentation in scientific practice that are key for scientific sense-making and articulates how engagement in these aspects happens both inter-mentally (between people) and intra-mentally (an individual's reasoning). Institutionally, peer review exerts critique on new knowledge claims in science and is comprised by a search for errors, which can be expressed as generation and evaluation of alternative possibilities that contrast with the new knowledge claim. Critique influences construction of claims, and this interplay motivates progress in sense-making. A classroom experiment is presented in which a high school physics class simulated the social interactions between authors and reviewers. These students' subsequent ability to engage in various forms of sense-making is contrasted with a control class that did the same activity but without simulating the relevant social interactions. The results suggest important support for a principled, practice-based way to simulate scientific discourse in classrooms to support sense-making.

For some time, science educators have been concerned that the way science is typically taught misses something fundamental and important to understand about science. Schwab (1962) articulated this in terms of rhetoric—although the way science is done is in reality a rhetoric of inquiry, typical science instruction presents a picture of science as a "rhetoric of conclusions." This is a concern because rhetoric, in the basic sense of communication among those working in scientific practice, is key to how scientific knowledge is produced. For the student, a focus on vocabulary and memorization, rather than on scientific sense-making, can result in "inert knowledge" (Whitehead, 1929). Remedying this motivates a considerable amount of scholarship in science education, including a focus on teaching ideas about science (e.g., Osborne, Collins, Ratcliffe, Millar & Duschl, 2003; Lederman, 1992), modeling (e.g., Stewart & Rudolph, 2001; Windschitl, Thompson, & Braaten, 2008), and argumentation (e.g. Berland & Reiser, 2009; Erduran & Jimenez, 2008) and is a central motivation for recent policy recommendations (Duschl, Schweingruber, & Shouse, 2007; Michaels, Shouse, & Schweingruber, 2008; National Research Council, 2012). A basic idea underlying these efforts is to teach science so that students learn to make sense of nature in ways that are fundamentally similar to scientific sense-making rather than merely paying lip service to facts and ideas.

Correspondence should be addressed to Michael Ford, University of Pittsburgh, 5525 Posvar Hall, Pittsburgh, PA 15260. E-mail: mjford@pitt.edu

This line of work has raised questions about what good scientific sense-making is and how it might be taught. One approach has been to focus on argumentation because it is through arguments that scientists make sense of nature. Instructionally, a popular approach has been to teach students about some general structure of scientific arguments. Structure has been characterized in a variety of ways, for example as the logic of controlling variables (e.g., Chen & Klahr, 1999); as a distinction between theory and evidence (e.g., Kuhn, 1991, 1993; Kuhn, Iordanou, Pease, & Wirkala, 2008); and as a set of components including claims, evidence, warrants, rebuttals, and the like (in terms of Toulmin's argument pattern; Toulmin 1958). Because such structural features are general, the idea is they can support scientific sense-making across the contextual details of different topics and natural phenomena.

The current article considers a different way of thinking about what students need to know in order to engage in scientific sense-making. The strategy here is to focus on students attaining a "grasp" of scientific practice—that is, an ability to participate in key forms of discourse and activity that form the epistemic basis of scientific claims. Although structural features of knowledge and arguments do reflect practice to an extent, what a student needs to know to participate in science appropriately goes beyond an understanding of structure. Engaging in scientific discourse and practice is dialogic, and the dialogue between voices that construct and critique ideas about nature in scientific ways is fundamental, not only to inter-mental argumentation but also to intra-mental (or internal, individual; Wertsch, 1991) sense-making. In this way, argumentation can be conceived as including key sense-making processes, both socially and individually.

Evidence from previous studies suggests this. For example, Ford and Kniff (2006; Ford, 2008b) suggested scientists and non-scientists made sense of scientific claims in the popular media in ways that are substantially different, and the differences reflect aspects of practice. Whereas scientists were more tentative and tended to critique claims as they figured out what they meant, non-scientists tended to accept these claims on their face. Also, whereas scientists tended to make sense of the claims by focusing on how data were collected and analyzed to consider what the claims meant, non-scientists were more likely to uncritically relate the claims anecdotally to personal experience. Rather than a dual process of positing meaning and checking its plausibility against the data, non-scientists made sense in a more singular process of folding meaning together with experience. Scientists seem to understand something about the discourse and practice of science that suggests for them the appropriate way to make sense of its products, which seems to be an interplay of constructing sense and critiquing it.

In another study (Ford, 2005), two classes of sixth graders experienced two weeks of instruction about experimentation. One condition focused on the control of variables strategy (CVS) as Klahr and colleagues have formulated and studied this structural aspect of scientific sense-making (Chen & Klahr, 1999; Toth, Klahr, & Chen, 2000). The other condition engaged students in a holistic, scaffolded simulation of key scientific practices such as constructing, critiquing, refining, and symbolizing measures. Assessments given following these two-week units suggested that whereas the CVS students were better able to evaluate comparisons as either confounded or unconfounded (that is, when variables were defined on a paper and pencil assessment), they were less able to construct a good experiment on an open, novel task. In contrast, students from the practice-oriented class did better at constructing a novel experiment. Their performances on both of these assessments were suggestive in that they seemed to reflect a duality of constructing and critiquing. When constructing a novel experiment, students tended to try tentatively a way of collecting data and then step back, reflect on it, critique it, and then revise their method of data collection. On the paper and pencil CVS assessment, they tended to look for flaws in the comparisons that were presented, even beyond merely checking to make sure settings on the defined variables were appropriate for the desired inference to be drawn.

The results of these studies tentatively suggest an essential duality in scientific sense-making that involves an interplay between constructing and critiquing. Structural accounts highlight components of scientific sense and rules of inference but largely do not attend to activity. The argument in what follows is the activity of scientific sense-making in practice is social and epistemically based on construction and critique, and sense-making for an individual is fundamentally similar. This raises a potentially important possibility—instruction that engages students socially in the interplay of construction and critique as they support scientific progress in sense-making may result in appropriation of these aspects of practice and an enhanced ability to make scientific sense individually as well. It may be that a "grasp of practice" (Ford, 2008a, 2008b; Ford & Forman, 2006) as such could be a key support for scientific sense-making not only for an immediate context, but also for subsequent contexts, essentially serving as a preparation for future learning.[1] This account of sense-making represents a possible specification of how scientific discourse and practice influence reasoning, an additional step beyond important work that has considered science and learning in terms of discourse (e.g., Lemke, 1990) and culture (e.g., Latour, 1999).

The present study further explores this account of scientific sense-making. Two high school physics classes conducted a two-week unit on ramp experiments. One class simulated the interplay between construction and critique and how it supports progress in sense-making. The other class did what is more common—students were precluded from construction and critique through provided instructions on how to do their experiments and through prohibitions on substantive critique or revision. The study sheds light on what students learned through analyses of interviews that followed instruction. The results imply that engagement in construction and critique helps prepare students to make scientific sense subsequently of novel scientific ideas. The evidence also reflects a duality in reasoning that the dialogic account of sense-making presented here suggests.

ARGUMENTATION, SENSE-MAKING, AND INSTRUCTION

A Focus on Argument Structure

Several lines of work about characterizing scientific reasoning and improving science instruction have emphasized structure of knowledge and arguments in one form or another, where "structure" can be thought of as a set of normative features of arguments or acts of arguing, general in form but instantiated in specific contexts. The basic idea underlying a focus on structure for learning is that students will learn how to argue and reason well by including these features in their thinking. For example, Klahr and colleagues (e.g., Chen & Klahr, 1999; Klahr, Chen, & Toth, 2001; Triona & Klahr, 2003) studied different ways of supporting students to learn the CVS. CVS is the logical fact that inferences can be made only from experimental comparisons that have

[1]Preparation for future learning is a specific phenomenon, described by Bransford and Schwartz (1999), that stems from observation of contrasting cases. Although I use the same words here, they are intended in a more general sense that what one learns in one context can influence one's ability to learn in a subsequent context.

particular features. In order to infer that a variable influences an outcome, the setting or measure of the variable of interest must differ across conditions, and all other relevant variables must be held constant. CVS is a structural description of the kinds of arguments that can be made about experimental results, as well as a structural heuristic that can guide design of good experiments.

Other psychological and developmental research on scientific reasoning also has emphasized structural features of arguments. A groundbreaking line of work by Deanna Kuhn and colleagues (e.g., Kuhn, 1991, 1993; Kuhn et al., 2008) has emphasized that sound arguments have a structural separation between a theoretical claim and the evidence for it. The importance of this separation has been described in contrast to a common tendency for people to believe what they want to believe, folding evidence into their existing ideas rather than viewing their ideas in a critical light. Rather than use one's idea only to make sense of a situation, one also can consider one's idea as an object for inspection itself. For this, a fundamental issue is having evidence that can be considered as independently bearing on the idea, in some sense. Kuhn (1991) articulated this perspective:

> In order to relate new evidence to their theories, people must know what their theories *are*. Only then can they examine evidence with regard to its bearing on the theory. If one thinks only *with*, or *through* a theory, rather than *about* it, one has little awareness of what the theory is—and what it is not. . . . As a description of "the way the world is," a theory offers little of value. It assumes power only to the extent that one can envision alternatives with which it competes, as well as evidence against which it can be evaluated. . . . (pp. 238, 267; emphasis in original)

Other approaches also have emphasized the role of evidence in argumentation. A considerably influential template used to scaffold students' construction of arguments is Toulmin's Argumentation Pattern (TAP). Several groups of researchers have drawn on TAP (e.g., Berland & Reiser, 2009; Duschl & Osbourne, 2002; Erduran, Simon, & Osbourne, 2004), which brings attention not only to claims and evidence but other structural features as well, such as warrants (a rationale for why evidence supports a claim), rebuttals (challenges to claims, evidence, or warrants), responses to rebuttals, and so on. There are several advantages to the TAP template, including that it is anchored in a broader theory in philosophy of science (Toulmin, 1958) and has a great degree of specificity.

The TAP template or some similar form of structure has guided the development of curricular materials and computational environments, under the idea that a structural template provides a clear set of expectations for the arguments students should construct, suggesting ways to think about what a good argument is in terms of its structural components. For example, WISE (Bell & Linn, 2000) and BGuILE (Reiser et al., 2001; Reiser, 2004) are computational environments that have been designed and studied for this kind of support. The general idea behind these important and successful efforts is that through prompting, or instructional scaffolding, students practice making claims about some natural phenomenon, typically in response to questions that are well designed for getting them to examine thoroughly the content and data that are provided about that phenomenon. Because this structural form of support is common across a good number of studies in science education, Bricker and Bell (2008) highlighted this category as scaffolding studies. Through such scaffolded experience constructing arguments, students learn not only the content but also reasons why the content is accepted as reliable by the scientific community—both the theoretical ideas of science, and its epistemic features as well (Sandoval & Reiser, 2004).

A Focus on Scientific Discourse and Practice

An alternative way of thinking about sense-making is through a focus on science's practices and patterns of discourse. Particular ways of acting and communicating are crucial for how science, as an institutional and cultural endeavor, makes sense of nature, and individuals who have a grasp of scientific practice are able to engage in scientific sense-making because they have appropriated these forms of discourse and practice.

Sense-Making Socially. Previous work has articulated a model of sense-making in scientific practice through cycles of construction and critique (Ford, 2008a, 2008b; Ford & Forman, 2006). From science studies the model draws the idea that scientific knowledge production is a communal endeavor and is a result of interactions among disciplinary peers (Longino, 2002; Pera, 1994). Individuals basically play two roles—constructors and critics of knowledge claims—within scientific communities, and progress in the construction of knowledge results from social interactions according to these.

To become scientific knowledge, a claim must be certified by the community of peers, and in order to be certified, it must pass a rigorous critique. Typically the community members that critique a knowledge claim are working both in the same area (if not on the very same problem) and in competition with the author of this knowledge claim (for grants, publication priority, prestige, etc.). Typically, knowledge claims are not immediately accepted, and critics provide specific reasons for rejecting their acceptance in light of the aims of the scientific endeavor, reasons that zero in on possible flaws in the warrants or evidence supporting the claim. The author of the claim then sets to work specifically on those flaws. When the flaws are removed and no more can be identified (and of course, the claim represents progress in understanding), the scientific community accepts the claim as certified scientific knowledge.

The notion of opposition is central to this sense-making process of iterative refinement. Critique can be defined as identifying errors, where the term "error" is intended here to mean that the state of affairs in nature is different in some way than is being claimed. For example, measurement error can be considered a category of this, if broadly we imagine that measurement error means some actual state of affairs in nature is different from reported instrument readings. Critique is an active search for ways that a claim might be wrong, which is about imagining how nature could be different than what is claimed, given the constraints imposed by the evidence and the information about it that is available.

This process of critique is fundamental in the sense that it strongly informs construction of claims (e.g., a claim is considered new knowledge only when peers cannot reasonably imagine any errors; a scientist constructs claims by anticipating critiques and ruling out errors). It is also fundamental in the sense that this process in scientific practice is responsible for the emergence of the structural features of science historically and institutionally. Mayo (1996) described the development of statistics in experimentation, for example, as a historical process of sense-making through assertion and critique in which potential errors became identified, measured, and otherwise managed. The structure provided by statistical experimental design is thus a *result* of critique historically. Institutionally, as identified possibilities for error became generalized from specific contexts (e.g., false positives are identified across multiple studies, then become coined in general as such), the statistical structure subsequently became normative.

The interaction of construction and critique in practice fundamentally underlies structure. Consider that no adherence to a structural description can guarantee sound scientific conclusions. A claim is not assuredly true because there is evidence to support it. Similarly, controlling variables is insufficient—there may be variables that influence an outcome that an experimenter is unaware of. For example, at one time it was considered absurd that a study participant's health could be influenced by his belief that he is receiving a medication. Yet from a structural perspective, the studies at that time were considered sound (controlling for relevant variables) even if they did not have a placebo condition. Of course, we now know there is a placebo effect and it is one of the influences on outcomes that has become normative to control. At any time in a scientific field's development, there are effects that are unknown and factors that influence things scientists are ignorant about. Progress and sense-making in any field of study is driven not merely by controlling for variables we know but by looking for influences on outcomes that we do not know.

Thus although structure can be an important guide, it cannot ensure sound results. Critique, although it cannot *ensure* sound results, is a key generative part of practice that stimulates progress and from which structural features of claims emerge—historically, institutionally, and even personally. The interaction between construction and critique, or proposition and opposition, is fundamental to how science works and how its practices support progress in sense-making. Therefore a "grasp" of practice thus conceived is a good candidate for what is important to understand about science (Ford, 2008a, 2008b; Ford & Forman, 2006), particularly if the field needs to orient itself toward what to teach students in contrast to memorization and vocabulary.

Sense-Making Individually. The notion that learning is dialogic has been well developed in other important lines of work (e.g., Matusov, 2009; Mercer & Littleton, 2007; Mortimer & Scott, 2003; Wells, 1999). Because sense-making in scientific practice stems from an interaction between construction and critique, sense-making for individuals reflects the same (Vygotsky, 1978). Bakhtin (1981, 1986) drew on the notion of "voice" to articulate what might happen mentally for individuals when their reasoning reflects patterns that exist in social interactions, but clearly without the literal presence of multiple individuals in one's mind.

In a Bakhtinian sense, sense-making is fundamentally oppositional. One can imagine the reasoning of a scientist when constructing an experiment or an argument about results as an interaction of voices. As the scientist makes sense of a phenomenon, not any sense will do. Because the claim that is being constructed will ultimately need to pass the critique of peers, and this critique will be a search for possible states of nature that contrast with the claim (and the evidence for it), then the success of the scientist depends upon her anticipating this, doing the same, and obtaining information that rules those alternatives out. So in a way, when she is making sense of a phenomenon, this process is informed by a critic's "voice," as well as by a constructor's "voice." This duality of diverse but complementary "voices" is a central feature of the reasoning that makes progress in sense-making, for the individual as well as for the community. It is also the case that the aim of the entire scientific endeavor, in a particular field or writ large (e.g., to explain natural phenomena), is an implicit feature of the claim. Not any sense will do, but some sense that represents progress in science, and sense that is reasonably shown (according to the standards of that field) to stand in contrast to alternative and demonstratedly unlikely possibilities.

The fundamental role of opposition goes beyond a scientist conducting inquiry to making sense of finished claims as well. Identifying meaning involves some consideration of what a claim is not, some awareness of the field or set of possibilities in which it sits, as well as consideration of what a claim is. In the inter-mental conversation in science, this is explicitly what the constructor of claims aims to do—rule out alternative possibilities, not only in the presentation and support of claims but also in the work that is involved in collecting data. The aim is to show that nature could be no other way. And, if meaning is defined by contrast, then consideration of contrast is involved in the process of understanding intra-mentally as well, for an individual making sense of that finished proposition.[2] For Bakhtin, understanding the meaning of an utterance involves meeting the voice of that utterance with appropriate "answering words" of one's own (Voloshinov, 1973, p. 102). If sense-making is relational, then it occurs when a proposition voice is met by a sense-making voice. This is not to say that the sense-maker necessarily entertains all the alternative possibilities that critics and constructors do. But it does mean that there is something fundamentally similar in these voices—a duality between active generation of definite possibilities regarding the state of affairs in nature and checking these against the knowledge claim being considered. Because this duality was instrumental in the claim's construction—in a way, this is what the claim is about—the way a learner makes sense of it needs to take this into account.

In both forms of sense-making, while constructing a new claim or understanding what a claim means, the interaction of "voices" is of a particular speech genre (Bakhtin, 1986). Because of this, learning to make sense scientifically involves becoming familiar with this genre, which involves understanding (a) what it means to construct scientific sense toward the community's aim; (b) what it means to critique a claim, evidence, and argument; (c) how these different voices interact appropriately in scientific practice; and (d) that this interaction supports iterative scientific progress. These things are distinct, yet they are part of a coherent whole. Having a grasp of practice thus involves understandings and abilities for each of these independently, but also an understanding of the whole, from which each ability gets its rationale.[3]

A focus on structure and scaffolding students to include particular features in their arguments is indeed one way of supporting their ability to participate in scientific practice. The possibility being put forward here, however, is that something fundamentally important gets missed in what students may take away from such experiences if critique is not included appropriately, and if the whole of how these diverse aspects of practice interact toward scientific progress is not represented with fidelity. Just as practice underlies structure historically, it also underlies the appearance of structure in activity and in reasoning. Under a structural view, because the focus is on constructing an argument, critique may be conceived merely as a means to this end, as a prompt for construction. It is typically not singled out for focus of instruction itself, and if the teacher remains the main critic, then students are not encouraged to learn how to critique—and

[2]Bransford and Schwartz (1999) focused specifically on how being exposed to contrasting features of something is related to a subsequent ability to learn, and therefore was dubbed "preparation for future learning." The argument here is similar in that contrast is involved, but that through a grasp of practice, one makes sense by generating contrasts, simply because this is the appropriate way to participate culturally in science.

[3]Holland, Lachicotte, Skinner, and Cain (1998) similarly noted the holistic character of the information involved in knowing how to negotiate one's identity within a cultural world. The case here is about what one knows about science, as a practice, in order to participate in it. I am grateful to Ellice Forman for pointing out this connection.

more importantly, how critique is generative for progress. Instruction under a practice view supports students to appropriate, in what they do and in what they think, a dualistic process of sense-making from which structural features emerge.

What Might be Evidence of a Grasp of Practice, or Oppositional Voice?

Earlier studies suggested that scientists and laypeople approach and reason about novel scientific claims differently (Ford & Kniff, 2006), and students that experience simulated practice as such seem better prepared to conduct their own inquiries (Ford, 2005). The particular ways that the more knowledgeable performers in both of these studies contrast with their counterparts suggest a grasp of practice. In both cases, the performances exhibited a duality in reasoning—a positing of some candidate and checking of it—that seemed instrumental for progress in sense-making. The scientists wanted to know more about how data were collected and analyzed so they could decide what to believe about a claim (Ford & Kniff, 2006; Ford, 2008b). The students in the classroom study made progress in their experiments by attempting a way of measuring, checking, realizing errors, and iteratively revising their apparatus (Ford, 2005).

The possibility being explored here is whether a grasp of practice can be taught in a principled way, even in a relatively short period of time, and whether the appropriation of constructor and critic "voices" can be detected through their specific implications. In what follows, I will refer to this more simply as appropriation of oppositional voice, which implies the duality and emphasizes that which is fundamentally generative toward sense-making. What implications in behaviors or performances would suggest students had appropriated an oppositional voice? Perhaps the clearest suggestion would be that if one has appropriated an oppositional voice, then one distinguishes between one's ideas and the evidence that bears on them (Kuhn, 1991, 1993), not because one has been prompted to do so, but because making sense as a participant in science involves *both* construction of sense and critique of it. The inclusion of critique necessarily pushes the sense being constructed into a different kind of thing—an idea that is not only used to look *through*, but also an idea that is looked *at* and scrutinized (Kuhn, 1991). Conversely, an oppositional voice might also discourage singular forms of sense-making that tend to fold all new information into conformity with one's favored ideas. Specifically, confirmation bias—the notion that one does not seek out disconfirming evidence—is a problem in reasoning that an oppositional voice would theoretically prevent or overcome (Klayman & Ha, 1987; Nickerson, 1998). Related to this is the notion that advanced reasoning exhibits an avoidance of jumping to conclusions, or what has been called "premature closure" (Chapman & McBride, 1992; Wollman, Eylon, & Lawson, 1979). Because sense-making includes the generative role of critique, this dualistic account provides a way of considering the reasoning that would avoid these problems.

The appropriate way to make sense of a claim in science is to test, to check, and this is a key implication here. Because this is a specific and appropriate move with its rationale in practice more broadly, an additional implication goes beyond that students *can reason* in these ways, but that they *do reason* in these ways appropriately (i.e., for scientific sense-making) without prompting when situations invite. Scientists know that it is appropriate, when making sense of a claim, to probe the evidence that is available for it, generating contrasts (Ford & Kniff, 2006; Ford, 2008). They do this simply because it is the appropriate way to react to a claim to understand it, in that situation. Thus in addition to the reasoning ability itself, another implication is noticing the kind of situation for which cuing these aspects of practice is appropriate. Evidence would be appeal to an oppositional voice, without prompting, in situations that invite sense-making.

In what follows, I describe a comparative study in which two high school physics classes engaged in ramp experiments. The strategy of the study was to control all features of instruction except those that simulate the interaction between construction and critique in scientific practice. The question that drives the study is whether, in assessment interviews afterward, there is evidence of an oppositional voice in the way students learn new content. Specifically, do students guard against premature closure (Chapman & McBride, 1992; Wollman et al., 1979) and confirmation bias (Nickerson, 1998)? Is there evidence that they distinguish between their belief of what a novel claim means and the evidence for it (Kuhn, 1991, 1993, 2005)? And is there evidence that students learned something particular to science: Do they scrutinize the claims in specifically scientific ways?

METHOD

Participants

This study was conducted in a working-class school district outside of Pittsburgh. Two classes were compared, both constituted by non-honors physics students in grades 10 and 11. Students in each class were administered a pretest for information regarding their comparability. Their scores on the test were not significantly different. The number of students in the treatment class was 16, and the number of students in the control class was 22.

Instruction

Both classes of students were co-instructed by their regular teacher and the author (denoted as "Instructor 1" and "Instructor 2" respectively in the transcripts below). Both classes experienced about 10 hours of instruction on a ramp experiment activity. Both classes were provided the same introduction and rationale of the ramp experiment unit: They were to learn about ramp motion, but also about experiments and how science works. They were to do experiments to answer the question, "How does steepness affect speed?"

Whereas one class ($n = 16$) experienced the *dual role* condition, the other class ($n = 22$) experienced a *standard lab* condition. The dual role condition was designed to engage students in the roles of constructor and critic, as they each do their part separately and together toward the goal of arriving at a scientific answer to a question. The central features of the dual-role condition are summarized in Table 1. The standard lab condition was designed to engage students in the same ramp experiments—to teach the same ideas about ramp motion and about the tools of scientific inquiry (e.g., graphs, distributions, patterns) that the dual role condition experienced—except without engaging students in construction or critique. In general, instruction led the control students to measure and graph the same ramp motion patterns that the dual role students did. It provided detailed instructions, supported students to draw the graphs, to include information about the range of their data in their graphs, to question what patterns the graphs might imply

TABLE 1
Key Features of Dual Role Condition

Materials	Rudimentary technology and materials - Wooden boards of defined lengths - Golf balls - Unit height boxes - Stopwatches - Whiteboards and markers Measurement error is salient with these materials and leads to efforts to limit and quantify it relative to systematic variation across conditions
Task	- Convince your peers that your group has the most scientific answer to the question: How does steepness affect speed - Two experiments with different operationalizations of steepness (a) hold board length constant and vary height (leads to a focus on pattern of change rather than change alone) (b) hold height constant and vary board length (leads to data that do not conform to expectation)
Activity structure	- Groups plan procedure - Groups take turns presenting and critiquing presenters - Groups revise procedures and execute iteration 1 of experiment 1 - Groups present and critique results of iteration 1 of experiment 1 - Groups revise procedure and execute iteration 2 of experiment 1 - Groups present and critique results of iteration 2 of experiment 1 - Groups execute experiment 2 (but do not discuss results)
Instructional guidance	- During group work, suggest ways to be more scientific - Prompt for audience to critique presenters - Model critiques and highlight what "more scientific" means in relation to a model critique

about ramp motion, and the like (the complete instructions are provided in Appendix A). However, rather than act as constructors (by letting students figure out what a scientific experimental set up and measurement protocol should be, for example) or critics (through prompts to critique their peers' and their own work), the students in the standard condition were led by the instructors to find the conclusions in their data.

The overall plan in both the dual role and the standard lab condition was to complete two versions of the ramp experiment. Each of these different ways of varying steepness leads conveniently to issues of what it means to be scientific. When ramp conditions incrementally increase height while holding plank length constant, the average speed (on the ramp) increases in large increments at first, but these increments get progressively smaller. Thus, each increment of height adds a progressively smaller addition in speed. This issue is convenient for focusing students on the rate of change, rather than just change itself—thereby serving as an example of science being more definite than everyday knowledge. That is, the scientific aim is to consider *how much* speed increases for each unit increase in steepness and in what pattern, rather than just *that* speed increases, which is the idea that students tend to begin with.

The second ramp experiment leads to another important and interesting issue. When steepness increases by using incrementally shorter planks from the same height, speed on the ramp does not increase, which is a counterintuitive result. That is, the time it takes for a ball to roll down any length ramp from a given height is directly proportional to the length of the ramp. In other words, the average speeds (distance divided by time) for all ramps of the same height are equal. This is a convenient issue because students tend to expect that in all cases of increased steepness, the measured speed will increase. This experiment thus poses a challenge for students to suspend this assumption and focus on the data, interpreting them in an analytically honest manner.

Within this context of ramp experiments and the issues that they conveniently shed light on, the strategy of the study was to engage two classes of students of similar ability in these, controlling for everything except a simulation of construction and critique as they unfold and work together in scientific practice. The dual role condition was designed to achieve this by iteratively asking students, in groups of three or four, to *construct* steps in the inquiry process. That is, on the first step students were given the question, "How does steepness affect speed?" and were asked to plan a ramp experiment with the materials (wooden planks of various length—2, 3, 4, 5, 6, and 7 feet, with four of the 4-foot planks available—boxes for resting the planks on, golf balls, and stopwatches). They were told that their task was to present the most scientific experiment plan for answering this question to their classmates, and the class together would decide which group's plan indeed was most scientific. During the presentations, the audience was prompted to critique the presenters' plans in terms of what ways they were insufficiently scientific, and the presenters answered their critics. Instruction explicitly scaffolded these conversations so they would be as scientific as possible (more on this in the next section below). After presentations, groups then decided whether and how to revise their plans and they moved on to conducting the experiments. A similar activity sequence was then repeated at several other stages—after groups had preliminary results, after they conducted a revised experiment, and in planning the second ramp experiment. Each time, the rationale was to come up with the most scientific answer to the question, present it to their peers, and persuade their peers to certify their answer.

The standard lab class did the same ramp experiments and included group presentations as well, but the rationale for presenters during these was not to persuade their peers that they had the most scientific answer, and the audience was not charged with critiquing the presenting groups. Rather, the purpose of the presentations was simply sharing results, as is the case in many secondary laboratory presentations.

Interviews

Immediately following conclusion of the unit in both classes, all students were interviewed individually. Interviews were conducted by the author and by a teaching intern that was observing these classes at the time. Each interviewed approximately half of the members of each class.

The interviews were designed to address three different content areas, in increasing "transfer distance" from the content related to the ramp experiments. The first content area was the second ramp experiment itself, the second content area was the law of free fall, and the third content area was represented by a brief popular science magazine article outlining what studies have shown about the dangers of phthalates (a chemical added to the plastic in baby toys) to the health of infants. The complete interview protocol is provided in Appendix B. Both interviewers agreed that for each question, the interviewer could ask a general clarifying question if necessary (e.g., "I don't understand, can you explain what you mean a little more?") and should provide one probe at the end, "Do you have anything to add?"

The rationale of the interview was to test whether the expected differences would appear in student responses. Specifically, dual role students were expected to learn and understand the content better than those in the standard lab condition as reflected in more "sustained consideration" (Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993); less premature closure (Chapman & McBride, 1992; Wollman, Eylon, & Lawson, 1979) and confirmation bias (Nickerson, 1998); greater distinction between novel claims and evidence for them (Kuhn, 1991, 1993, 2005); and greater scrutiny of claims in specifically scientific ways.

Ramp Experiment Content Area. The first questions in the interview addressed results from the second ramp experiment that the students had conducted. (Recall that students were asked to draw conclusions from their group results individually, not with peers.) Students were provided with two graphs—one that represented the relationship between plank length and time of roll on the ramp, and another that represented a transformation of these results as plank length versus average speeds (plank length divided by each roll time). The data were represented as "box and whiskers" plots, which corresponded to the way students in both conditions were encouraged to represent their results in both the second iteration of the first experiment and the second experiment. The graphs shown to students in the interview appear in Figure 1.

Students were asked two questions about these graphs:

1. The group that did this claimed these results prove that steepness affects speed. If you were in the audience, how would you critique this claim?
2. What do you think these results show about the relationship between steepness and speed?

The aim of the ramp experiment content area was to detect differences in students' ability to critique ramp experiment results and to make the appropriate sense of these results.

Interpreting graphs is an important skill, and in this case the graphs represent an invitation to check one's assumptions. Students typically assume that steeper ramps will result in faster speeds, because this is given from everyday experience. However, when steepness is operationalized with constant heights and varying board lengths, the average speed on all ramps is the same, a fact originally discovered by Galileo. The results reflected in the graphs confirm this fact, because

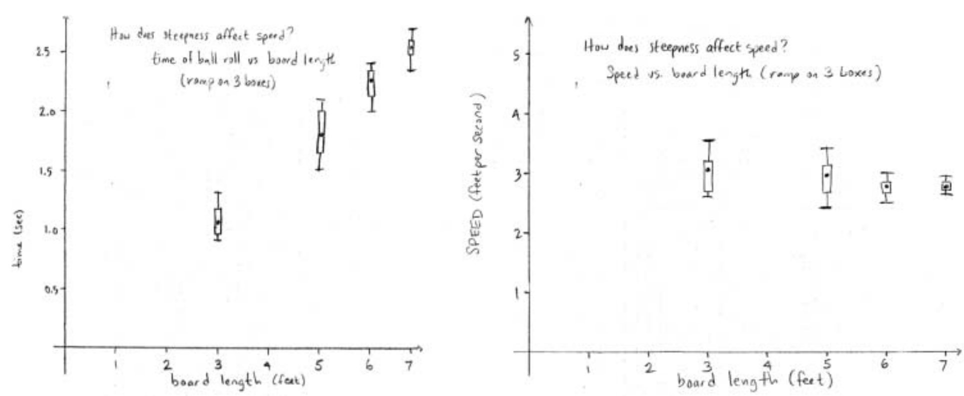


FIGURE 1 Ramp experiment data provided in the interview.

in the board length versus time graph, the measurements are best modeled linearly from the origin. Correspondingly, in the board length versus speed graph, the results are best modeled as a horizontal line. However, there is ambiguity in the graphs because of measurement error. Therefore, drawing this conclusion from the graphs requires one to challenge a preconception that steeper ramps always result in greater speeds. As such, this task is an invitation for oppositional voice. An oppositional voice would prevent premature closure, would check confirmation bias, and would be reflected in a distinction between the interpretation students were constructing and the information provided in the graphs. Students who had attained a grasp of practice would check their assumptions, and those who had not would fold the results into their preconceived idea that steeper means faster.

The Law of Free Fall Content Area. The second content area during the interview was the law of free fall. Because this study was conducted in the third and fourth weeks of the physics course, students had not yet learned this idea. The law of free fall is a mathematical pattern that was a key part of Galileo's theory of motion. He asserted that an object accelerating regularly from rest over equal time intervals (no matter what the size of these intervals) travels, in those time intervals, successive distances that are as the odd numbers starting with one. That is, during the first time interval the object will move some distance, then in the second interval it will move three times that distance. In the third interval it will move five times that distance, in the fourth seven times, and so on. The contemporary version of this law is expressed not as distances moved in each time interval, but the total distance moved. In this way, the equation *total distance* = *(constant)*time*^{2} results in a pattern of total distances as successive squares (1, 4, 9, 16, 25 . . .). This law holds not only for free fall but also for motion on inclined planes (with minimal or no resistance), as Galileo first demonstrated it.

Students were presented with the law of free fall in tabular form, as shown in Figure 2, and students were told that this table represented the pattern that Galileo discovered by which objects fall. Students were asked to figure out the pattern in this "law of free fall" and to explain it to the

time (Seconds)	total distance dropped (feet)	differences	difference multiples
0	0		
1	16.08	16.08 ÷ 16.08 =	1
2	64.30	48.22 ÷ 16.08 =	3
3	144.69	80.39 ÷ 16.08 =	5
4	257.22	112.53 ÷ 16.08 =	7
5	401.90	144.68 ÷ 16.08 =	9

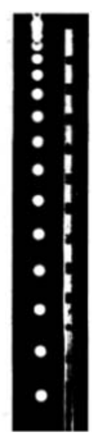


FIGURE 2 The law of free fall representation.

interviewer. The rationale behind this was to detect any ways in which the two ramp experiment conditions had prepared students differently to learn a novel pattern of motion, even though students in both conditions had been thinking about, measuring, and analyzing *different* patterns of motion for the previous two weeks.

This task is an invitation for oppositional voice, with information that is different from the ramp experiment the students had conducted. An oppositional voice would be reflected in responses that would generate a possible pattern that is reflected in the table and then would check whether that pattern obtains in it. In contrast, students who had not appropriated an oppositional voice would likely offer a pattern that they believed the table reflected and would see *through* that idea, merely using it to make sense without checking it.

Poisonous Teething Rings Content Area. For the third content area, students were asked to read a brief article, "Alarm Sounds Over Toxic Teething Rings," that appeared in a popular science magazine, *The New Scientist* (July 14, 1997). The full article appears in Appendix C. This article has been used in previous research to explore differences in the ways students and scientists interpret media reports of scientific studies (Ratcliffe, 1999). It is readable for high school students (categorized as "average readability" by Flesch formula; Ratcliffe, 1999, p. 1089) and contains important details about the methods by which conclusions have been reached. Thus it has the features that scientifically literate people should pay attention to and evaluate as they interpret the article's claims and decide how it impacts their lives.

This article highlights the possibility that potentially poisonous chemicals, called phthalates, which are added to plastics to make them soft enough for babies to chew on, seep out of teething rings and are ingested. It provides details about studies that tested various brands of teething rings by shaking them in artificial saliva and then measuring the quantity of phthalates that were transferred to it. It provides three sample measurements, 2,219 micrograms (noted to represent more than 44 times the amount allowed under European Union food laws), 1,044 micrograms, and 9 micrograms. In addition to these and other details, it is suggested that any baby that sucked on these teething rings for three hours would ingest these quantities of phthalates, and that this would harm internal organs.

After reading the article, students were told that another student believes that "this study *proves* these teething rings are hurting babies" (with emphasis as indicated). Students were asked whether they agreed or disagreed, and they were allowed to elaborate. Afterward, students were asked why someone might have the contrary opinion (i.e., if they said they agreed, why might someone disagree and vice versa).

This interview strategy was designed to explore how student experiences in the dual role condition influenced their reasoning about novel scientific content. It was inspired by Kuhn's (1991) strategy to invite students to reason about both sides of an argument (a strategy subsequently also employed by Means and Voss, 1996). This is an ideal instrument for measuring appropriation of oppositional voice because it has been used before multiple times, and it meshes theoretically with what is posited here about appropriation of constructive and critique voices. One would expect oppositional voice to appear not only through critiquing the argument presented in the article, but also through a critical attention to the appropriate details in the article, those that characterize a scientific approach to the question, in order to measure one's confidence or skepticism in it. In contrast, students who had not appropriated an oppositional voice would be susceptible to believing the claim that the plastic toys are hurting babies without critical scrutiny of the scientific details of the studies reported in the article. That is, whereas presence of an oppositional voice would motivate one to entertain an idea as an object for scrutiny, absence of it would motivate one to see *through* an idea, using it only to organize sense rather than to check the idea itself, prematurely closing the issue.

Data Collection and Analysis

Both instructional conditions were video recorded with a single camera, which was placed in the back of the room and focused forward, for capturing whole class discussions. These data were transcribed and analyzed for themes, in terms of characterizing descriptively overall how instruction in each condition unfolded.

All interviews were audio recorded and these records served as data. The author listened to the recorded interviews and developed a coding scheme that indicated theoretically relevant ways that the dual role students' answers differed from the standard lab students' answers. Appendix D contains the coding scheme itself along with sample student utterances that exemplify the qualitative codes.

To measure coding reliability, a second coder who was not familiar with the coding rationale was trained on 10% of the of the data corpus. Training involved coding the data with the information provided in the coding scheme; then the author provided feedback regarding the codes that were initially assigned and explanations for these using quotes from the interviews. Following this, the second coder coded 50% of the corpus. Coding results from this were compared to the initial code assignments and inter-rater reliability was calculated at 91%.

RESULTS

Description of Instruction

This section describes the way instruction in both conditions unfolded. Although the key features of the instructional conditions were described previously, the results of these efforts appear here. This distinction could be considered in terms of Remillard's (2005) distinction between a planned and an enacted curriculum. This detail is intended to serve two functions: (a) to show how the dual role condition unfolded (how it represented construction and critique) and to make clear how it contrasted with the control condition; (b) to highlight ways in which the social interactions of construction and critique were supported instructionally and engaged by students. That is, students needed to be taught what to do to engage with each other so that construction and critique in scientific practice were simulated, and they seemed to learn this to some extent, as some uptake of these roles is evident. The detail is intended to show what aspects of this students did without trouble, what aspects seemed more difficult, and how instruction supported those aspects that were more challenging.

Dual Role Condition. Students in this condition were posed the question, "How does steepness affect speed?" They were provided relatively unstructured materials for answering this question, as described above. They were organized into groups of three or four and asked, in their groups, to develop a plan for an experiment that would answer the question. After they did so, groups presented their plans to the rest of the class. Groups illustrated aspects of their plans on whiteboards.

At first, students basically felt the answer to the question was obvious, that steepness will increase speed. They also felt that they did not need to pay attention to details of data collection or analysis. During the first presentations of results, the two instructors prompted students in the audience to ask questions of the presenting groups and point out ways that their answer to the question could be more scientific. Students generally did not do this, so the instructors modeled such critiques. Because of this criticism, students then realized that they had not understood what was meant by scientific—in this case, to have data that reflect a small range (within conditions) and little measurement error, and to have a rational argument about the pattern of speed change as the steepness increases.

Students felt that the instructor critiques were not fair because they had not been given explicit instructions for how to do the experiment. This affect motivated the instructors to draw explicit distinctions between following detailed directions, which is what students are typically expected to do in school, and the gradual but consistent work scientists do to zero in on a phenomenon with solid data and specific patterns. It is often the case that scientists need iteratively to revise their experiments as they learn about their apparatus and the phenomenon under study.

Students agreed that they could get a better scientific answer in light of what they had learned (both about what it means to be scientific and about the measurement challenges on the ramp apparatus) and set out on a second attempt. The presentations of results in this second round were much more clear and specific, and student critiques focused both on data collection and on claims about the posited pattern. The instructors challenged students to consider not only averages (which they had graphed exclusively) but also the range of measurements for each condition. The unit ended with the students doing the second version of the ramp experiment, in which ramps of different steepness paradoxically yield the same average speeds. Although student groups carried out this experiment and individuals recorded their conclusions, presentations were not made. What students concluded from the results of this experiment was probed and recorded in an interview question following the unit's end (described below).

To reiterate, at first, the unit unfolded with instructors prompting and modeling critique. When it became apparent that students did not understand what "scientific" meant in the context of their experiment, the instructors specifically challenged students to critique ways in which their and their peers' experiments could be more scientific, while also pointing out what scientific means in this context. In this way, the modeled critiques of the instructors consistently "upped the ante" in service of scientific critique. Initially, students presented results that were not clear, specific, or concerned with error. Critiques upped the ante by pointing out these areas for improvement. Then when students produced more specific and reasoned arguments about patterns, the instructors upped the ante again by skeptically posing the possibility that the pattern may be different if ranges of data were considered rather than just averages. As may be expected, students' ability to critique was thus related to their ability to construct good experiments, and both were inseparably bound to (and limited by) their knowledge of scientific practice. The teaching of critique was through the instructors' voices, exhibiting what was unscientific about the current answer and evidence. These voices were framed explicitly in terms of what students should do (and did) for their colleagues' and their own work, iteratively, to make their answer and evidence as well checked as possible through consideration of alternatives and specific enough to be checked through prediction and test.

Following are sample transcripts to give a sense of how these issues played out. The students both went along with the unit's development and resisted. The resistance can be traced, it seems, to violated expectations and a sense that critiquing and being critiqued are out of place in a high school classroom (despite the instructors' explicit declarations to the contrary), where the expectations are rather that critique should be limited to student failure to follow directions.

The first transcript is from approximately the third hour of instruction. Students had completed their first iteration of data collection and were presenting their data.

Instructor 2:	And so I'm asking now—let's just look at it. Is it indeed scientific?
Female 1:	No.
Instructor 2:	How is it not scientific? Why do you say that?
Female 1:	It's not exactly the same.
Instructor 2:	It's not the same? Really?
Female 1:	Uh-huh.
Instructor 2:	Okay, remember everybody in the audience, your job during this unit is to point out things that aren't scientific, right? I'm just showing you what your job is. You need to be looking at this and ask yourself, "Okay, is this—do I buy this? Am I ready to say this is going to light up my house?"
Male 1:	They didn't use 25 degrees.
Male 2[4]:	They didn't need to.
Female 2:	We didn't have time.
Male 3:	What (Ben)?
Male 4:	I think maybe one of you missed the—didn't use the protractor right because at 45 degrees you got 1.14. This one has .63.
Female 3:	We realize that.

In terms of instructional support, this excerpt shows how the prompts to critique were combined with a challenge to consider whether the results were scientific enough. In terms of student activity, one student said that if their data collection had been scientific, they should all get the same answer. This was a problem for this student because there were considerable discrepancies in the results, which led to explicit consideration of the way they had collected data. Thus, students did seem to have ideas about what it means to be scientific, whether or not their ideas were in line with what these features are.

The second excerpt, from later in the same class period, reflects the way students were thinking about error. Instructor 1 was challenging one group, which posited that the pattern was a line, despite the fact that one of the points (an average of the trials in that condition) was considerably distant from the line. A male student says that there is probably a measurement error responsible

[4]This convention denotes whether the speaker is male or female, and whether he or she is different from other speakers in the same excerpt. Thus "Male 1" is not necessarily the same speaker that is labeled "Male 1" in the following excerpt. However, instructor labeling is consistent: "Instructor 1" is the regular classroom teacher and "Instructor 2" is the author. All student proper names are pseudonyms.

for that, so there really is no problem. The instructor pushed them to consider why other points should not be ignored as well.

Male 1: So say someone does screw up in timing it right; he didn't click it fast enough or whatever, that takes care of it really.
Instructor 1: So this one [gesturing to a data point] is the one that got screwed up, right? We should get rid of that one?
Male 1: Pretty much.
Instructor 1: Because, you know, you have that error or whatever. And we know this one was right, though, right? So we know—does everyone know why you're keeping this one and getting rid of this one?

{Pause}

Male 2: Wait. I'm confused.
Instructor 1: Well I'm just saying . . .
Male 3: We're not really knowing. This is just a day-to-day perspective. We're not really going by that . . .
Instructor 1: By that data?
Male 3: No. It's all an average.
Instructor 1: But shouldn't we kind of go by the data? Isn't that what it . . .
Male 1: Well we average it . . .

This episode suggests more about what the students considered scientific. The students acknowledge that there is error in a data point that does not agree with their posited pattern. Error was declared post hoc—that is, after they realized it did not fit the line. In this way, error seemed to function for students as a rationalization of why the discrepancy should not motivate a reconsideration of that pattern. The students seemed to hold two contradictory ideas. On the one hand their scientific argument should be based on the data. On the other hand, their argument should not be held accountable to the data. Indeed, the transcript suggests that their notion of "average" may be instrumental in allowing their data to simultaneously function in both of these ways. Instructionally, the episode reflects how the instructor was pushing students to reconsider this position.

Shortly thereafter, at the end of this class period, the students expressed frustration at what the instructors were asking them to do.

Instructor 2: No what I'm saying is when you get problems with numbers that we were pointing out, you know, we're not convinced that you have a scientific answer.
Male 1: So why did we do this project?
Instructor 1: We're trying to . . .

{Crosstalk}

Female 1: Yes.
Male 2: So everything we did so far is wrong?
Instructor 1: What's that? So the question is why are we bothering to do this?
Male 2: So is everything we did wrong?
Female 1: The whole world is backwards, okay?

Instructor 1: Well I think you had some good ideas, but I don't think that—I mean, I saw every single group interpret this different. I saw people holding protractors like this, like this, like this, like this. So I think what it means to be scientific is really what we're after here. And this is kind of like the question we're going to use to help us understand that better. So that's really what we're doing.

Male 3: That's because the directions weren't specific.

This excerpt not only provides an example of the energy and frustration of the students, but also that they felt that the instructors had led them into a trap. We had not told them how to do the experiment, and then we told them that they were wrong. This violated their expectations about "fair play" in school, in which students are held accountable only for those things the teacher had provided directions for. Asking the students to hold their work to a scientific standard, when "scientific" was not completely clear to the students up front, was perceived as unfair.

At the beginning of the next class, the instructors tried to make clear to the students (a) that they were not expected to have a perfect answer to the question on the first try, (b) that they were expected to keep working iteratively to come to the most scientific answer they could, and (c) that they should contrast what they know about school with what scientists do.

Instructor 1: Got you. All right. So the last thing that we did is we were presenting our work to each other. And that kind of gets uncomfortable. We understand that. Right?

Female 1: No, we were all done. You were just telling us how we we're always wrong.

{Laughter}

Instructor 2: Now the one thing about school is school basically trains you to be just like—you go out on first down and you missed the ball and you say, "Oh I missed. I didn't do it right," and then you go and sit down. That's what school does. Because tests are designed that way. Tests are always sort of like, "Okay you got one shot here." If you miss the shot then you fail the test and you're, you know, that's it; it's on your record. You don't get a chance to go and fix it. You don't get a chance to keep plugging away and keep plugging away and keep plugging away.

So what we're trying to do here is shift the game here a little bit in school to be more like what scientists do. Yes, scientists are wrong. And they keep plugging away. It's not about being the immaculate reception. It's about being willing to keep plugging away. It's about working. You don't need to be a genius. What you need to be is somebody with a backbone and to have—you know, be willing to get in there and get your hands dirty and figure stuff out.

Does that make sense?

Male 1: Yes.

Instructor 2: So, and in a way, yes, we were trying to say this isn't good enough. And you guys all know it's not good enough.

Male 2: I thought it was pretty good.

Instructor 2: You told me—I mean you told [Instructor 1] it wasn't good enough and you told me it wasn't good enough as I was going around to your groups. So you know, I'm not telling you anything new. What's new, though, is that you probably might not have expected that we are asking you to keep plugging away.

The final excerpt is provided to illustrate some of the progress students made after this point. After re-collecting data a second time, they engaged in a conversation about measurement error, critiqued each other's work and conclusions, and argued for and against various possible patterns (a linear relationship; speed doubling for every two boxes added; an inverse relationship). At the point of this transcript, the class seems split on the latter two possibilities. At the end, one student seems to have confounded the notion of pattern and a linear relationship. Nevertheless, the excerpt reflects the engagement of the students: They had thought through the reasons for positing one pattern or another and were considering issues in defining the pattern in terms of rate of change.

Male 1: All right. I think that we've found an answer to the question, because two groups that thought completely different got the same answer.
Instructor 2: What is the answer?
Male 1: There's an inverse relation.
Female 1: There's a (inaudible).
Male 1: They have the same graph as us—it's the same answer.
Male 2: I think it doubles.
Instructor 2: And you're convinced?
Male 1: Yes, I am a hundred percent convinced that that is the answer.
Instructor 2: Okay.

{Crosstalk}

Instructor 1: Hey, we have some side conversations going on (inaudible).
Instructor 2: (Bonnie) Is yours (inaudible)?
Female 2: What it says is, like, if it was any other kind of relationship but an inverse relationship, you would see a pattern go up the same way, because math and slope and what not, but an inverse relationship, you can't find. Like, if you try and find the slope of a curved line, you have to use tangents, and they'll be different at every point. It won't—so, you weren't averaging your answers and going to be a double or anything like that, they'll just all—they're all different.
Instructor 1: Are you, just to clarify, are you saying you can't really find a pattern for a curved line?
Female 2: Yeah.

After the first experiment was concluded, the class then spent the final period and a half collecting data on the second experiment (varying steepness by holding height constant and varying plank length) and analyzing results. Groups were told that they would not present their results but would be evaluated individually. Groups worked on their analyses for one class period, after which the unit was concluded and the students were asked to record their individual "most scientific" argument for their answer. These records were not collected but rather served to help students organize their thoughts for the interview, as the first interview question was to explain (when presented with actual data) what their conclusion was regarding this paradoxical experiment.

Standard Lab Condition. The standard lab condition conducted the same two ramp experiments, but students were not asked to think about either how to construct the experiment or how to critique others' experiments and results. In order to preclude student thinking about how to construct the experiments, detailed instructions were provided on how to set up the apparatus, how to collect data from the ball rolls, how many trials to run, how to tabulate and graph these trials, and the like (see Appendix A). If students had questions about the directions, the instructors demonstrated what to do, for example, how to graph the data by putting height on the x-axis, time on the y-axis, and how to draw ranges. Like the *dual role* condition, these students also conducted two experiments: one by varying height, and one by varying plank length. The *standard lab* students presented their results to their peers and reached the same point as the *dual role* students did—comparing results in terms of their reliability and positing patterns that would best represent the way speed changes on the ramp. However, unlike the *dual role* students, the *standard lab* students were not asked to critique.

Key themes were treated in the *standard lab* condition as well. During the first iteration, instructors prompted students to think about whether they would do anything differently, given the results they had and the difficulty of concluding a pattern of speed change with certainty. (The first iteration for the standard lab group also evidenced a large degree of within-condition variation, which was clear because the instructions directed them to graph both the average and range for each condition.) Finally, the *standard lab* group also addressed the distinction between systematic and random error in their data. The instructions directed that two different people should measure the ball roll times and that both these sets of results should appear on the same graph in different colors. In several cases, one color of results was consistently above the other color of results on the graph. The following excerpt is from a conversation about this issue, illustrating that the *standard lab* students were engaged in the data and how influences on their data might be traced to physical aspects of how they collected them.

Male 1: The steeper the ramp is, the less time it takes for the ball to roll. And it seems to be, like, a—the graph seems to be a curve. So if you add more boxes, the times will change less and less.

{Pause}

Instructor 2: Any questions?

{Applause}

Instructor 2: Excellent job. Any questions?
Female 1: Any questions? You always have questions.
Instructor 1: I'm still sucking in and I was taking care of logistics, so I'm looking.
Female 2: Heath has a question.
Male 2: What are the lines? I can't really understand what you did.
Female 3: What lines? These ones?
Male 1: Yes, like . . .
Female 1: That's the range. That's like, this was our highest, like, the longest it took the ball to roll. And this is the shortest it took, and then the average is in the middle.
Male 2: Oh, I understand now.
Instructor 2: So if you were going to say whether the red pair or the blue pair was more careful in their data collection, who would you say?
Female 4: The blue.

Female 3:	The blue one because the ranges are smaller, so that means there's probably less error in the time.
Male 1:	I didn't do it.
Instructor 1:	You did the blue, James?
Male 3:	I did the orange.
Instructor 1:	Oh all right.
Instructor 2:	So what do you think happened there with . . .
Female 4:	This one?
Instructor 2:	Yes.
Male 3:	That I kind of just went one, two, three, threw the ball and it just kind of rolled down the ramp and bounced.

The standard lab condition was designed to mimic as much as possible what was taught in the dual role condition, but without students generating ways of *constructing* the experiment and conclusions from it, and without students being prompted and supported to *critique* their peers' experiments, analyses, and conclusions in terms of them being scientific. In this excerpt, students were dealing with systematic and random sources of error and relating them to specific aspects of data collection. However, this did not arise or function as critique.

Standard lab students also conducted the second experiment by varying plank length, and, like the dual role students, were told that they would not present their results but would analyze the results in their group and would be evaluated individually. Like the dual role students, the standard lab students also then finished the unit by figuring out individually what to conclude from the second ramp experiment results, which they then were asked about in the interview.

Interview Results

The interview results will be presented in three sections, according to the three content areas. Overall, dual role student responses evidenced some telling differences with those of the standard lab class. Dual role students critiqued the ramp experiment claim better and were better able to draw correct inferences from the ramp experiment results. They had more productive interactions with the law of free fall table and tended to be more skeptical of the poisonous teething ring claim. Although these content areas were different and students were asked different things about them, overall dual role students were more actively inquisitive about the content, gave it more sustained attention (Resnick et al., 1993), and avoided premature closure (Chapman & McBride, 1992). Most interestingly, the dual role students' way of interacting with the novel content seemed to be supported by their generation and entertainment of more possibilities regarding what was being asserted by these content claims.

Ramp Experiment Content Area. As noted above, the second ramp experiment represented an outcome that was counterintuitive to what students typically believe, that is, that in all cases, steeper ramps result in faster speeds on those ramps. The interview was a way to evaluate what students had learned from the unit in terms of being able to infer a valid conclusion from data, even if that conclusion is counterintuitive.

Number and kind of critiques mentioned. Not surprisingly, when prompted to critique the claim that "these results show that steepness affects speed," students in the dual role class provided on average more critiques than the students in the standard lab class. Dual role students on average mentioned 1.5 critiques, whereas standard lab students mentioned on average 0.455 critiques. A two-tailed t-test on these results was statistically significant ($p < .001$). In addition, one quarter of the dual role students (4 of 16) not only provided critiques of the inference in light of the evidence but also requested more information about how the data were collected that appeared in the graphs. That is, dual role students not only appealed to the evidence that was provided but also considered that the way those data were collected might impact the confidence they had in them. No standard lab students requested information beyond the graphs. Thus, it seems that the dual role students critiqued not only the conclusion drawn in light of the results but considered ways of critiquing the results themselves in ways that they had done during the ramp unit, by appealing to the way the ball roll events were set up and how data were gleaned from them.

Interpreting data correctly. Dual role students were also better able to interpret the results of the second ramp experiment correctly. In this situation, this is evidence that they were better able to resist jumping to a conclusion that, in all cases, steeper inclines result in greater speeds—because as noted above, in this case, it does not. The time it takes for a ball to roll down any ramp length from the same height is directly proportional to the length of that ramp. Thus, average speeds on these ramps are equal.

Student responses were coded into four categories of increasing levels of sophistication that reflect oppositional voice:

1. The student replies not only agreement with the claim that "steepness affects speed," but also adds that steeper is faster. This code would reflect students who were uncritical of their assumption, who prematurely closed the issue, interpreting the results merely *through* this idea.
2. During instruction some students in both conditions mentioned that the experiment was not a fair way to compare speeds because such a comparison can be made only when ball roll lengths are equal. The second category notes students who seemed bothered by this issue. For example, one student stated, "I don't know, the lengths are different, it doesn't make sense."
3. Students who said the results implied that the average speed is the same across conditions. This code reflects oppositional voice in that the issue was not prematurely closed, but the assumption that "steeper is faster" was checked against the results.
4. Students who said that the speeds are the same and also noted that the different patterns reflected in the two graphs were compatible yet different ways of displaying the same ramp motion results. Results of this coding are displayed in Figure 3.

These results suggest that the percentages of students categorized into each level differed across conditions (Mann-Whitney U, $p = .029$). First, whereas almost half of the dual role students (7/16) interpreted the results correctly, thereby resisting premature closure and confirmation bias, only 3 out of 22 of the standard lab students did so. Second, the difficulty that standard lab students had interpreting the results correctly seemed to stem from the fact that they were more

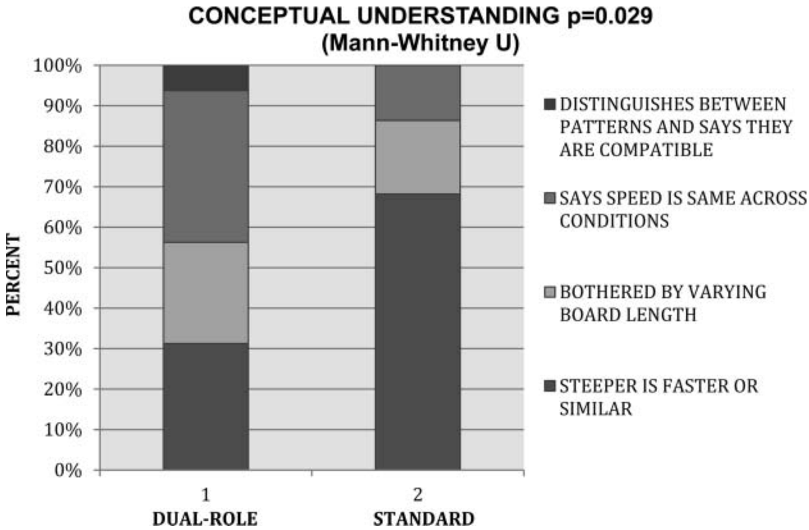


FIGURE 3 Percentage of students coded into categories of understanding ramp experiment results.

likely to hold on to the preconception that "steeper is faster." Whereas only about 30% of dual role students interpreted the results as "steeper is faster," almost 70% of the standard labs did so. Recall that students were looking at two graphs (Figure 1). There were two aspects of these graphs that students generally pointed out to justify their conclusion that "steeper is faster." One was simply that the board length versus time graph displayed a positive slope, and students seem to have interpreted this as more speed (despite the fact that longer boards correspond to less steepness). The other was that the board length versus speed graph suggests longer board lengths may have slightly smaller speeds (despite the fact that distributions overlap). Thus it seems like the standard lab group had difficulty interpreting the results correctly because they allowed their preconception that steeper is faster to bias their interpretation of the data.

Taken together, these two aspects of the results suggest that what distinguished the dual role from the standard lab group's reasoning was that the former resisted premature closure and confirmation bias and questioned the data in ways that are in line with oppositional voice.

Law of Free Fall Content Area. Following the ramp experiment, students were then shown a high speed multiple exposure photo of a ball falling and a table of numbers that represents the law of free fall. Although very few students (one from each class) were able to understand accurately the law of free fall from this somewhat cryptic table, there were some interesting differences in what was visible in their reasoning. Because this context is akin to encountering a new and difficult idea for the first time, it is a good situation for evoking a process of questioning and scrutinizing something new. Of particular interest here was evidence of this questioning and scrutiny. The coding scheme that was chosen to illustrate differences between the student responses simplified the questioning to mention of numbers and evidence that students were distinguishing a possibility of what the table might mean from the actual numbers in the table. The attention to the relevant information and what students did with it are the salient features

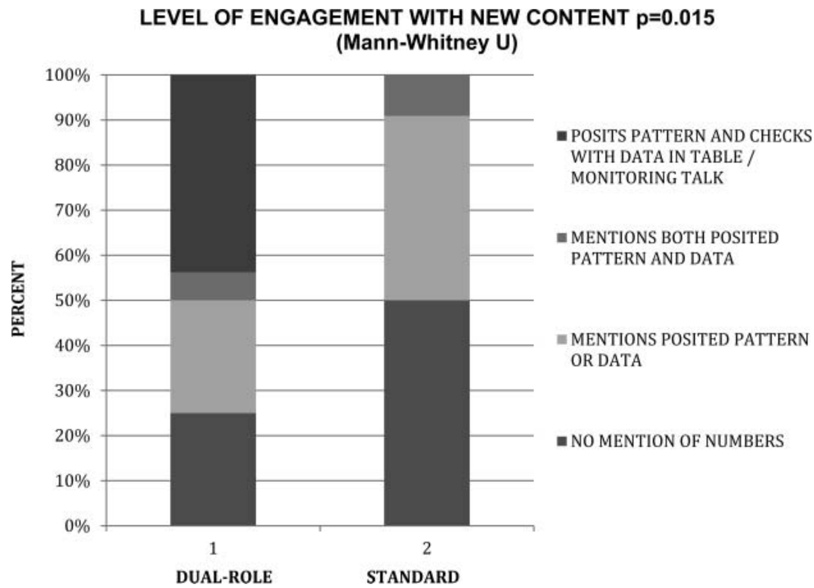


FIGURE 4 Categorization of students' responses to the law of free fall table.

here that reflect oppositional voice. Students who had appropriated oppositional voice engaged a reasoning pattern that held a possible interpretation as an object for inspection, for relation to the information in the table.Figure 4 represents the results of this coding, as well as the fact that the difference in percentages of students coded into each category was statistically different across the two classes (Mann-Whitney U, $p = .015$).

Two things are striking about this way of coding student responses. Whereas almost half of the students in the dual role class posited a pattern and checked the numbers in the table to evaluate whether this pattern was correct, none of the students in the standard lab did this. Positing and checking can be considered not only a coordination of theory and evidence (Kuhn, 1991, 1993) but also a reflection of propositional and oppositional voices as they function in scientific practice. This contrast between how the two classes of students responded is more stark if we consider that half of the students in the standard lab class did not mention numbers at all—neither by positing some pattern in the table nor by mentioning numbers that appeared in the table. Only 2 of the 22 students in the standard lab condition mentioned both a possible pattern and mentioned a number in the table, and there was no evidence that these students were trying to relate the two to each other, at least not aloud. As such, this fact suggests that the overall function of oppositional voice—stimulating a process of scrutiny, questioning, and evaluating—was less apparent in the standard lab responses than in the dual role responses.

Poisonous Teething Rings Content Area. Dual role students were more critical than the standard lab students of the claim that teething rings are hurting babies. Analyses of student responses to the interview questions about the article suggest that there were important differences not only in levels of skepticism but also in the ways students from the two classes went about making sense of what the information provided in the article might mean.

Whereas a strong majority of students in the standard lab group (18 of 22) were in unqualified agreement with the fictional students' claim that "this study *proves* that teething rings are poisoning babies," only about 30% of the dual role students were (t-test, $p < .01$). This result meshes with the differences in how students from the two classes evaluated the ramp experiment data in the first content area. For standard lab students, a driving idea served to organize and interpret the information provided, as is the case with confirmation bias. In this case, it may be that, like the notion of "steeper is faster," the idea that teething rings are poisoning babies organized and perhaps even filtered information, helping students interpret the provided details or assign significance to them, contributing to premature closure. The smaller proportion of dual role students in agreement with the fictional student's assertion suggests that the idea did not function in the same way for them.

More can be inferred about differences in the students' reasoning about the article through considering critiques. The dual role students mentioned more critiques than the standard lab group on average (t-test, $p < .05$). Whereas dual role students mentioned almost 2.5 critiques on average, standard lab students mentioned just below 1.5 critiques on average.

More interesting is the function these critiques seemed to serve for the dual role students. Specifically, they seemed to function as part of the sense-making process. Consider two ways of categorizing the critiques that were mentioned. One way of organizing them is in terms of when they were mentioned. Recall that there were two interview questions on this content area: First, students were asked whether they agreed with the claim that "this study *proves* that teething rings are hurting babies." After students responded to that and provided reasons for their opinion, the next question asked students to consider the contrary opinion. Although the number of critiques mentioned *after* this second question was not significantly different, the number of critiques *before* it was statistically significant ($p < .01$).

For the standard lab students, initial reaction typically was agreement that the study proved that the teething rings were poisoning babies. Thus, the question about the contrary opinion was, "Why might someone think that the study does *not* prove teething rings are hurting babies?" For most standard lab students, this second question was a prompt to critique the studies and inferences in the article. When the critiques were categorized as to whether they were mentioned before or after this prompt, it is perhaps not surprising that most of the standard lab student critiques were mentioned after this prompt. In fact, only 6 critiques total were mentioned by standard lab students *before* that prompt, resulting in an average of about 0.27 critiques per student. Whereas the number of critiques mentioned after the prompt were not significantly different between groups, the number of critiques mentioned before the prompt were (t-test; $p < .01$).

In contrast, most of the dual role student critiques were mentioned in response to the *initial* question, when they were considering whether they believed the article proved that teething rings were hurting babies. That is, for them, critique was something they found related to making sense of what the article meant, rather than merely imagining what a skeptic might be thinking. This is evidence of oppositional voice in the sense that it is evidence of scrutiny and questioning while the student is in the process of deciding what the claim means and what confidence it should be accorded. It is also suggestive that what distinguishes the standard lab students in contrast to this is premature closure.

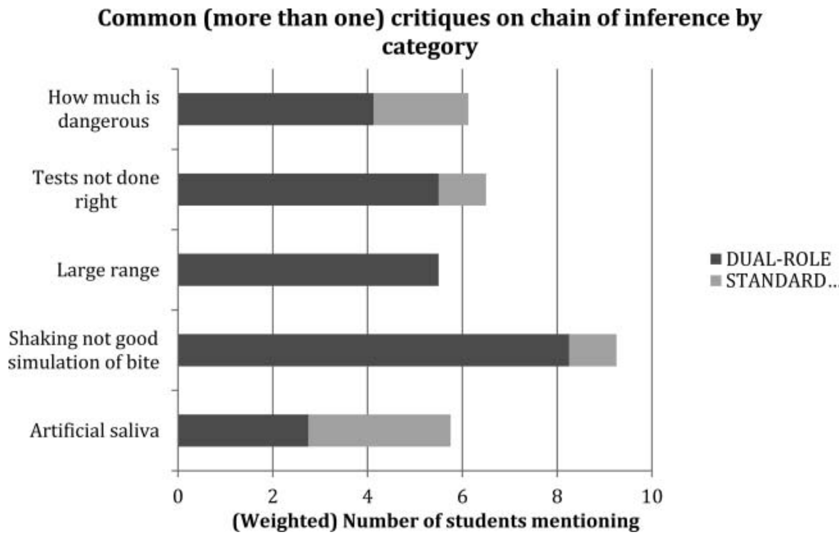


FIGURE 5 Number of students from each class mentioning specific relevant critiques.

Since the article is about a *scientific* question, additional insight into dual role students' reasoning can be gleaned from another way of categorizing the critiques. Although scrutiny and questioning are important in gleaning meaning in many domains, in science it is important to ask the right kinds of questions and pay attention to the right details. For this context, critiques were separated in terms of whether or not they concerned scientific aspects of the studies or their conclusions. Critiques were categorized as "on the chain of inference" or "idiosyncratic." Whereas the former are critiques that point to something in the article, idiosyncratic critiques are simply not about information in the article. Examples of these critiques include things like, "Maybe someone was angry at the company and purposely put the poison on the teething rings after they left the factory," and "Aliens did it."

When the critiques were categorized this way, the statistically significant difference (t-test; $p < .01$) between the dual role students and the standard lab students was in terms of the *article-relevant* critiques, not the idiosyncratic ones. Both groups mentioned an average of just over one idiosyncratic critique per student, but whereas the dual role group mentioned an average of almost 1.5 article-related critiques, the standard lab students averaged just over 0.5 article-related critiques. Figure 5 provides a look at the most common relevant critiques and the (equally scaled) number of students that mentioned each in the two classes. This way of considering student responses to the baby toy article suggests that dual role students were not only scrutinizing and questioning the article, but crucially that they were paying attention more often to the information that is relevant, *scientifically*, to the meaning of the claim and judging confidence in it. That is, dual role students seemed more in line with the "conversation," or "speech genre" of science—and an oppositional voice in science seems to be evident in their consideration of this other artifact of scientific practice.

When posed with a claim and evidence for it, dual role students seemed to ask themselves spontaneously, "as opposed to what?" That is, they seemed to have considered the claim that the teething rings were poisoning babies as one among multiple possibilities. Moreover, this consideration of multiple possibilities extended to states of affairs described in the study itself. For example, an alternative possibility was that when babies chew teething rings, phthalates are not released, but during three hours of submersion and shaking in artificial saliva, they are. Similarly, artificial saliva may be composed of some substance that invites phthalates to be secreted, or is less lubricating than real saliva, or there might be other differences that matter. And more obviously, these all support consideration of the alternative conclusion—that the studies do *not* prove the teething rings are poisoning babies.

DISCUSSION

Although it is not surprising that dual role students were better able to critique the ramp experiment results, it is interesting that they found the appropriateness of critique to extend beyond that content area to completely different content areas. Moreover, it is intriguing that the function of critique seemed to be to support sustained attention (Resnick et al., 1993) to the content encountered, avoiding premature closure (Chapman & McBride, 1992). In contrast, the standard lab students tended to take the meaning of the content they encountered on its face. They tended to interpret the content *through* their ideas, assuming that steeper is faster and that teething rings are poisoning babies, for example. The dual role students tended more to look *at* their ideas, examining them in light of the information provided (Kuhn, 1991).

An ability to avoid premature closure and give sustained attention to the content likely has both domain general and domain specific aspects to it. One might argue that the differences between the two groups in this study stem not from appropriation of a scientific oppositional voice but from some general differences like students being aware that they were supposed to think hard about the questions. Indeed, this makes sense because the dual role students experienced much more encouragement (through critiques of their work) to be careful and think things through than did the standard lab students. Other general aspects of difference are possible as well.

However, there are specific aspects of *scientific* practice that are reflected in the differences in student reasoning reflected in the results, and these aspects are precisely those that were emphasized in the dual role condition. Science aims for definite and specific claims, and for evidential support, often in the form of data that are themselves constructed, in the sense of being collected and represented and summarized in some posited pattern. In this practice, the oppositional voice motivates consideration of alternative possibilities on each of these levels: data collection, posited pattern, and conclusion. Most evidently in the teething rings question but also in the ramp experiment question, dual role students generated alternative possibilities on these levels.

If we consider the dual role condition, we can draw reasonable connections to how the oppositional voice in scientific practice was learned. First, it seems particularly relevant that instruction focused on the fact that "scientific" answers are those that are highly defined and specific, rather than vague, and those that are well-supported with evidence, which is often in the form of data. Thus students who grasped this tended more to look for specificity in content, scrutinizing and questioning it, in light of other specific alternative possibilities.

Second, dual role students were explicitly asked to identify potential errors in their and their classmates' results and inferences. Identifying errors involves generation of alternative possibilities that caused the results to be what they are, possibilities other than the actual state of affairs regarding rolling balls. By critiquing their classmates' results, they apparently became not only better able to make their own results and conclusions more rigorous but also developed an awareness that results and conclusions are constructed generally through careful attention to ruling out errors and alternative possibilities. Moreover, this action has its rationale and function in the broader aim of scientific practice, and it was framed as such in the instruction.

Third, it seems important that the dual role students faced the difficulty of getting good data on how steepness affects speed. They had to figure out how to minimize within-condition variation by dealing with details regarding how the ball was released, how the stopwatch was operated, and the like. In these efforts we might consider the actual time it takes for a ball to roll down the ramp as one possibility, and time measurements that deviate from that as alternative possibilities. Through grappling with the task of finding out what the actual time is, with only the ability to generate measurements at their disposal, students faced the reality that the actual time is one among many possibilities. What distinguishes the actual time from the others is errors—or, in other words, causes that influence the generation of a time measurement that are unrelated to the ball's roll (e.g., sticky fingers, a tilted plank, slow reflexes). In this way, dual role students spent considerable time attending to the field of possibilities at this level of detail and likely developed a respect for the challenge it is to distinguish the actual rolling time from the others.

Perhaps because of such practical difficulties, students in the dual-role condition seemed more likely to focus on the relevant details during the interviews in all three subject matter areas. These details were relevant for productive use of uncertainty in sense making in each of them. In the ramp experiment, uncertainty was reflected in the distributions. The dual role students were more likely to identify this and draw implications for the kinds of patterns that seemed most likely. The standard lab students, in contrast, tended toward certainty rather than uncertainty in that they saw what they assumed and were sure about it. For the law of free fall, dual role students seemed uncertain about what the pattern might be but also were able to posit and check candidates in a productive way. The standard lab students did not tend to engage this dual process. And with the baby toy article, dual role students seemed to acknowledge uncertainty in the results, pointing out more often that they were not consistent, and that there may be problems with what was done to generate them. It is important to note that this skepticism was not counterproductive to sense-making, as one can imagine unabated skepticism might be. Rather, it was skepticism that prevented premature closure and seemed to support progress in sense-making productively—because its role in sense-making was to aid in its construction.

There are important differences between constructing a scientific claim for critique by a community, as the dual role students had done during instruction, and making sense of a finished scientific claim as they did in the interviews. However, there are important similarities in these activities, and oppositional voice is offered here as a possible explanation of why experience with one would affect ability in the other. As a practice-based theory, the idea is that students had learned holistically how scientific sense is supported through a productive interaction between construction and critique. Oppositional voice is just what one does with a claim, because it is such a central aspect of practice. Because finished claims are a result of this practice, an implication might be that oppositional voice is relevant for gleaning meaning from it as well. In a way, this is an account regarding "what transfers," although a practice perspective is quite different from a traditional transfer account about abstract structure and application to novel contexts.What transfers in this account is oppositional voice, but clearly other holistic information about the practice of which it is a part is essential as well. If oppositional voice is a good explanation for the results of this study, then transfer might be considered in terms of something learned in one mode of participation in a practice (i.e., arguing about claims under construction) supporting abilities in a different but related mode of participation in the same practice (i.e., making sense of novel finished claims). In each, oppositional voice is relevant to sense-making, albeit in different ways. In the former, it's relevant for generating possible errors so that uncertainty can motivate progress. In the latter, it's relevant for being aware that the claim sits among a set of alternative possibilities and that the evidence about it is relevant for identifying how it contrasts with those. As part of a practice, oppositional voice has its rationale in each for supporting sense-making. As an appropriation of practice, students learned not only what oppositional voice is about but under what conditions and why it is useful for sense-making in these different but related modes of participation in science.

Although the oppositional voice explanation is coherent with the data and the instruction, there are other ways of attributing what students learned. For example, rather than oppositional voice, it could be that students learned to play the roles of constructor and critic appropriately. They may also be said to have mastered thinking strategies, much the way that these are conceived and urged to be taught (e.g., Kuhn, 2005).

These data do not distinguish among competing explanations for the effects noted. However, they are suggestive of a phenomenon that may be considerably important for several reasons. First, the duality of construction and critique clearly motivates and supports progress in scientific practice. This duality likely is reflected in scientific reasoning as well, and the account here fundamentally shifts the focus away from a singular emphasis on sense-making in terms of construction. Its also emphasizes scientific critique as something important for students to learn in itself and how such disciplinary skepticism can motivate progress in sense-making. Because of the fundamental shift this account represents, it may provide a way of framing productive research programs on learning to reason scientifically. Second, the notion that scientific practice is an interaction of construction and critique can be a useful way to orchestrate classroom discussions. It is convenient because, for teachers, this way of modeling activity emphasizes the kinds of social interactions students should be taught to have with one another and why these can be productive if done properly. It is important to have ways of moving key discourse processes into classrooms especially in light of recent science education policy documents (Duschl et al., 2007; Michaels et al., 2008; National Research Council, 2012).

There are some important connections between the perspective offered here and other approaches to supporting student learning in science, and future work should further explore these. For example, Engle and Conant (2002) characterized support for "productive disciplinary engagement" in terms of problematizing some aspect of nature for inspection, providing student authority to make sense of phenomena, and ensuring there is accountability and resources necessary for making sense. The perspective here may help further theorize the relationship between authority and accountability—in specifically disciplinary terms. Because critique so strongly informs construction of sense in science, it seems that disciplinary accountability needs to be taught in order for students to understand how to achieve disciplinary authority in the sense that they make (cf., Forman & Ford, in review). This of course implies a notion of authority beyond merely allowing students to make their own sense of phenomena. It implies disciplinary authority that is achieved and is based on community roles of constructor and critic and how these interact in practice. It therefore highlights the importance of disciplinary norms being supported in the classroom community (Driver, Newton & Osborne, 2000).

The aim here is to highlight the possible ways that conceiving of reasoning as a dual process, patterned after scientific practice, could be useful. If reasoning is shaped by practice, then it is theoretically reasonable that an oppositional voice can have a constructive role in meaning-making (cf., Billig, 1996). It is also empirically plausible in that students can, in a relatively short period of time, come to understand the productive role that uncertainty can play in constructing scientific knowledge claims and in making sense. Without such an oppositional voice, meaning-making is less a search for a specific candidate within a field of possibilities and more of an imposition of sense through adherence to a privileged idea that might remain vague and unchallenged.

ACKNOWLEDGMENTS

The research reported in the article was supported by a National Academy of Education/Spencer Postdoctoral Fellowship.

REFERENCES

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin, TX: University of Texas Press.

Bakhtin, M. M. (1986). *Speech genres and other late essays*. (C. Emerson & M. Holquist, Eds.; V. W. McGee, Trans.). Austin, TX: University of Texas Press.

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.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26–55.

Billig, M. (1996). *Arguing and thinking: A rhetorical approach to social psychology* (2nd ed.). Cambridge, England: Cambridge University Press.

Bransford, J., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61–100.

Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning science and their implications for the practices of science education. *Science Education*, *92*(3), 473–498.

Chapman, M., & McBride, M. L. (1992). The education of reason: Cognitive conflict and its role in intellectual development. In C. U. Shantz & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 36–89). Cambridge, England: Cambridge University Press.

Chen, Z., & Klahr, D., (1999) All other things being equal: Children's acquisition of the control of variables strategy, *Child Development*, *70*(5), 1098–1120.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.

Duschl, R. A., Schweingruber, H. A., Shouse, A. W., National Research Council (U.S.). Committee on Science Learning Kindergarten Through Eighth Grade., National Research Council (U.S.). Board on Science Education., & National Research Council (U.S.). (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483.

Erduran, S., & Jimenez-Aleixandre, M. (2008). *Argumentation in science education: Perspectives from classroom-based research*. Dordrecht, the Netherlands: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, *88*, 915–933.

Ford, M. J. (2005). The game, the pieces, and the players: Generative resources from alternative portrayals of experimentation. *Journal of the Learning Sciences*, *14*(4), 449–487.

Ford, M. J. (2008a). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*, 404–423.

Ford, M. J. (2008b). "Grasp of practice" as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, *17*(2&3), 147–177.

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, *30*, 1–32.

Ford, M. J., & Kniff, K. J. (2006). Groundwork for progress supporting scientific literacy: An expert-novice study. Poster presented at the Annual Meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Forman, E. A., & Ford, M. J. (in review). *Authority and accountability in light of disciplinary practices in science*. Manuscript submitted for publication.

Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S. M., & Tzou, C. T. (2002). Science inquiry in a digital world: Possibilities for making thinking visible. In H. V. Oostendorp (Ed.), *Cognition in a digital world* (pp. 253–284). Mahwah, NJ: Erlbaum.

Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.

Klahr, D., Chen, Z., & Toth, E. E. (2001). Cognitive development and science education: Ships passing in the night or beacons of mutual illumination? In S. M. Carver and D. Klahr (Eds.), *Cognition and Instruction: 25 years of progress*. Mahwah, NJ: Erlbaum.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228.

Kuhn, D. (1991). *The skills of argument.* New York, NY: Cambridge University Press.

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

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*, 435–451.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press: Cambridge, MA.

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. (1990). *Talking science: Language, learning and values*. Westport, CT: Ablex Publishers.

Longino, H. (2002). *The fate of knowledge*. Princeton, NJ: Princeton University Press.

Matusov, E. (2009). *Journey into dialogic pedagogy*. Hauppauge, NY: Nova Publishers.

Mayo, D. (1996). *Error and the growth of experimental knowledge.* Chicago, IL: University of Chicago Press.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, *14*, 139–179.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking*. New York, NY: Routledge.

Michaels, S., Shouse, A., & Schweingruber, H. (2008). *Ready, set, science! Putting research to work in K–8 science classrooms*. Washington, DC: National Academies Press.

Mortimer, E., & Scott, P. (2003). *Meaning making in secondary classrooms*. Maidenhead, UK: Open University Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, cross-cutting concepts and core ideas*. Washington, DC: National Academies Press.

Nickerson, R. S. (1998). Confirmation bias: An ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas about science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, *40*(7), 692–720.

Pera, M. (1994). *The discourses of science*. Chicago, IL: The University of Chicago Press.

Ratcliffe, M. (1999). Evaluation of abilities in interpreting media reports of scientific research. *International Journal of Science Education*, *21*(10), 1085–1099.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.

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. KIahr (Eds.), *Cognition and instruction: Twenty-five years of progress*. Mahwah, NJ: Erlbaum.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246.

Rouse, J. (1996). *Engaging science: How to understand its practices philosophically*. Ithaca, NY: Cornell University Press.
Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic supports for science inquiry. *Science Education*, *88*, 345–372.
Schwab, J. J. (1962). The teaching of science as enquiry. In J. Schwab & P. Brandwein (Eds.), *The teaching of science* (pp. 1–103). Cambridge, MA: Harvard University Press.
Stewart, J., & Rudolph, J. L. (2001). Considering the nature of scientific problems when designing science curricula. *Science Education*, *85*, 207–222.
Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, *18*(4), 423–459.
Toulmin, S. (1958). *The uses of argument*. New York, NY: Cambridge University Press.
Triona, L. M., & Klahr, D. (2003) Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments *Cognition and Instruction*, *21*, 149–173.
Voloshinov, V. N. (1973). *Marxism and the philosophy of language*. Cambridge, MA: Harvard University Press.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. New York, NY: Cambridge University Press.
Wertsch, J. V. (1991). *Voices of the mind*. Cambridge, MA: Harvard University Press.
Whitehead, A. N. (1929). *The aims of education and other essays*. New York: The Free Press.
Windschitl, M., Thompson, J., & Braaten, M. (2008). How novice teachers appropriate epistemic discourses around model-based inquiry for use in classrooms. *Cognition and Instruction*, *26*, 310–378.
Wollman, W., Eylon, B. S., & Lawson, A. E. (1979). Acceptance of lack of closure: Is it an index of advanced reasoning? *Child Development*, *50*(3), 656–665.

APPENDIX A: "STANDARD LAB" INSTRUCTIONS

RAMP EXPERIMENT 1

Materials:
Wooden plank of length 4 ft.
Priority mail boxes
Golf ball
Stopwatch

What we're doing:
Measure time *on* the ramp, then use that to calculate speed. Each group varies the height of the ramp, and completes 4 conditions.

How to construct the ramp:

– Place 4-ft board on box(es), for four conditions: 2, 3, 4, 5 boxes.
– Make sure board is about 1 inch overlapping from the edge of the box, and that this overlap is the same for each ramp condition. (You may want to use the line on the box near the edge to keep this consistent.)

How to roll the ball:

– place the ball near the top edge of the ramp, and hold it with one finger.
– Release the ball by lifting your finger

How to measure the time:

- Practice a few times to get the hang of it before doing the real data collection!
- The person who times the ball roll should not be the same person who releases it.
- Position yourself so you can clearly see when the ball leaves the end of the ramp.
- Consider placing a box at the end of the ramp and stop the timer when the ball hits the box.
- The timer person should slowly say, “1, 2, 3,” and on 3 the releaser lifts her finger from the ball, and at the same time, the timer starts the stopwatch.
- Press the button when the ball reaches the end of the ramp.
- Do ten trials for each condition with one timer and releaser; do ten more trials for each condition using different people in each position.

How to record your data:

- use the table provided for recording the twenty trials for each of four conditions.
- Make the first ten trials one set of people, and the last ten trials the other set of people.
- **How to analyze your data**
- First see which team got better data: The better data are more likely to be the set that has the smallest range. The range is the difference between the largest and smallest number. Record the range for each in your table.
- Calculate the average for each team’s measurements. So, for the first ten trials, take the average. Then take a separate average of the second ten trials. Record these on your table.
- Using graph paper, plot these two averages for each of the conditions. Also, with a different color, plot the ranges of each. Put the number of boxes used on the X-axis, and the times on the y-axis. Make sure to plan your scale on the Y-axis so that all the times will fit.

How to draw conclusions

- Which team had better results? Are both results just as good?
- Given what you see on the graph, what is the pattern by which steepness affects speed? Is there a different answer for different measurement teams?
- Try to go beyond “steeper means faster” to be as specific as possible. Try to define the pattern, as if you’d be able to predict conditions that are not on your graph (extrapolation).

	2 BOXES	3 BOXES	4 BOXES	5 BOXES
1				
2				
3				
4				
5				
6				
7				
8				
9				

(Continued on next page)

(Continued)

10
RANGE
AVERAGE
1
2
3
4
5
6
7
8
9
10
RANGE AVERAGE

RAMP EXPERIMENT 2

How to construct the ramps:

– Rather than varying steepness by varying the number of boxes, or the *height* of the ramp, this time vary the steepness by varying the *length* of the ramp while holding the height constant.
– Use 3 boxes for each ramp.
– Make 4 ramps using the 3ft board, 5ft board, 6 ft board and 7 ft board.
– Make sure board is about 1 inch overlapping from the edge of the box, and that this overlap is the same for each ramp condition. (You may want to use the line on the box near the edge to keep this consistent.)

How to roll the ball (all directions same as in experiment 1):

– Place the ball near the top edge of the ramp, and hold it with one finger.
– Release the ball by lifting your finger

How to measure the time (all directions same as in experiment 1):

– The person who times the ball roll should not be the same person who releases it.
– Position yourself so you can clearly see when the ball leaves the end of the ramp.
– Consider placing a box at the end of the ramp and stop the timer when the ball hits the box.
– The timer person should slowly say, “1, 2, 3,” and on 3 the releaser lifts her finger from the ball, and at the same time, the timer starts the stopwatch.
– Press the end timer button (same as start button) when the ball reaches the end of the ramp.
– Do ten trials for each condition with one timer and releaser; do ten more trials for each condition using different people in each position.

How to record your data:

– Make a table for recording the twenty trials for each of four conditions.
– Make the first ten trials one set of people, and the last ten trials the other set of people.

How to analyze your data

– First see which team got better data: The better data are more likely to be the set that has the smallest range. The range is the difference between the largest and smallest number. Record the range for each in your table.
– Calculate the average for each team's measurements. So, for the first ten trials, take the average. Then take a separate average of the second ten trials. Record these on your table.
– Using graph paper, plot these two averages for each of the conditions. Also, with a different color, plot the ranges of each. Put the board length used on the X-axis, and the *speeds* on the y-axis. **Note: you will need to calculate speeds from the times by dividing the board length by the times. Only do this for the averages and the largest and smallest values, as in the second part of experiment 1.**

APPENDIX B: INTERVIEW PROTOCOL

Ask the student's name and what period they have physics.

Part A

Provide the two graphs of ramp experiment results, one time vs. board length and the other speed vs. board length. Ask the following questions:

(1) The group that did this claimed these results prove that steepness affects speed. If you were in the audience, how would you critique this claim?
(2) What do you think these results show about the relationship between steepness and speed?

Part B

How fast do things fall? Here is a picture of a ball in free-fall, taken using multiple flashes and a high speed camera.

Galileo claimed that things speed up during free fall according to a simple pattern. Here is a table that represents this pattern. (Show the table)

(3) Describe the pattern.
(4) Pretend that I don't know what the pattern is, and it is your job to explain to me what the pattern is and how to see it in this table. Teach me how Galileo said objects fall.

Part C

Ask the student to read the teething ring article silently.

Tell the student that a student from a different class said this study *proves* (with emphasis) that these teething rings are

(5) Do you agree or disagree with this student?
(6) Explain your position (why or why not).
(7) What might someone argue that believes the same thing as this student?
(8) What might someone argue who disagrees?

APPENDIX C: ALARM SOUNDS OVER TOXIC TEETHING RINGS. (*The New Scientist,* July 14, 1997)

STORES in Denmark, Sweden, Italy and Spain are taking plastic baby toys off their shelves after Danish scientists revealed that some teething rings release large amounts of toxins called phthalates. Denmark is calling for strict European Union limits for phthalates and other chemicals in toys.

The Danish Environmental Protection Agency (DEPA) studied 11 makes of teething rings - soft plastic toys that babies chew to relieve the pain of emerging teeth. Three of the rings, made by the Italian company Chicco, one of the world's biggest makers of toys and childcare products, released large amounts of phthalates when shaken in artificial saliva.

Lisbet Seedorff, head of DEPA's chemicals division, says a baby who chewed a Chicco Softy Vinyl Sweets ring for three hours would ingest 2219 micrograms of phthalates per kilogram of body weight. This is 44 times the maximum amount permitted in food under EU law. The second worst teething ring released 1044 micrograms of phthalates in three hours. The third released just 9 micrograms, but DEPA still recommended that it should be taken off the market as a precaution because a baby's bite would squeeze out more phthalate than the shakers used in the tests.

Phthalates are used to soften PVC plastic. Those used in the teething rings have previously been reported as damaging to the liver and reproductive system, and capable of causing cancer, says Seedorff.

Danish stores and KF, Sweden's largest retail store, agreed to take the teething rings off their shelves at the end of May. In addition, KF sent 75 other soft plastic toys to be tested for phthalates. This month, stores in Italy and Spain also removed the Chicco teething rings from their shelves. But they remain on sale elsewhere, as Chicco has not taken them off the market. The company could not be reached for comment.

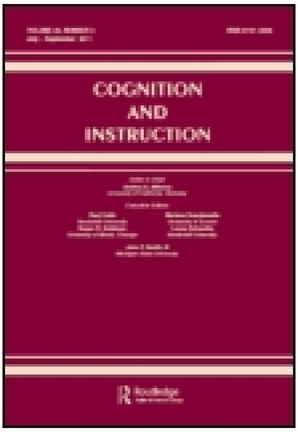
Last week, the trade association Toy Industries of Europe set up a crisis committee to look into the Danish study. "We'll see if we agree with how the tests were done," says Maurits Bruggink, head of the association. "We wonder if you should apply a legal limit meant for food to toys."

Seedorff says the main problem is that the EU's directive on toy safety limits the release of heavy metals but sets no standards for other chemicals, stating only that toys must not be dangerous. Denmark and Sweden are pressing for changes to the directive. The Danish study has been sent to EU scientific advisory committees that are expected to propose limits for chemicals released by toys later this year.

APPENDIX D
Coding Scheme and Sample Student Responses

Ramp Experiment Results		*Definitions and Example Responses For Each Code*
Critique?	No critique	"Don't know," assume steeper is faster, the results make sense or partly make sense
	example	"I don't know."
	Some critique	Critiques the work that led to the results in any way
	example	"I don't think that would be right, because if you look at the graph, they're the same."
	Considers data collection	Goes beyond the graph to how the data were collected, either by recommending what else the people could do or by asking for more information about how they did it
	example	"They should try some more tests with more boxes and see if they get the same results."
Number of critiques	count	
Understanding	Steeper is faster or similar	When asked "what do these results say about how steepness affects speed," Says steeper is faster or something similarly shallow or indecipherable.
	example	"Um, through time and board length."
	Bothered by varying board length	Bothered by two variables changing (board lengths not held constant so nonsensical to compare speeds on these ramps) or doesn't make sense because two graphs have different results.
	example	"I like this one [the board length vs. time graph] because it shows an incline of speed, I don't know. But this one [the board length vs. speed graph] doesn't change much, so I don't know. I never quite understood this."
	Says speed is same across conditions	Says speed does not change
	example	"I don't, I don't think it, that it would be right just because speedness, er steepness would affect speed if you look at this graph because those two are the same, to have them just about the same except for the ranges, the dots are the same."
	Distinguishes between patterns and says they are compatible	Clearly distinguishes between two displays (speed and time) and understands they are compatible
	example	"Well then this one shows that it relatively went the same speed the entire time, which I can understand because we proved that. Um, and so this one, well this one makes sense then too because as your board gets longer, the more time it's going to take the ball to roll down, so, they are understandable."

Galileo's pattern	No mention of numbers	no mention of numbers during the answer
	Mentions posited pattern or data	Mentions either numbers in the table or a posited pattern in numerical form (e.g., "it doubles")
	example	"Um, it increased by two? Like the difference. It's linear, it's going up, it's positive. . . . It increased relatively about a hundred each time, then it kinda gets bigger up there."
	Mentions both posited pattern and data	Mentions both numbers in the table and a posited pattern in numerical form
	example	"For each second, you start out with 1, you get the 16.08, then you divide that by 16.08, so I guess the 16.08 is the feet that it's dropped, the distance, and then you would get the 1, and then you would go to 2 seconds, and you would take the distance again which is 64.30 and divide it by 16.08, you get three. For the third time in seconds which is 3, the distance is 144.69, you do the same thing again, divide by 16.08 and you get 5. And you keep going."
	Posits pattern and checks with data in table/Monitoring talk	Either posits a pattern and checks it with the numbers in the chart or clear meta-talk (e.g., I'm thinking here that. . . . ")
	example	"It's going up 2 every time, like between this one, it's the sixt—, er the 48, it's going up basically like 48, times two or—I don't know how to explain it. This one's 16 times two, then 48 times two—oh, that's not right. . . . It's roughly two between each. It's basically all."
Teething rings article		
Unqualified agreement?	y/n	agrees with no qualifications/does not agree or has qualifications
Number of critiques	count	
Number before/after prompt	count	
Number on/off info in article	count	



Cognition and Instruction

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hcgi20

A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning

Michael J. Ford [a]

[a] University of Pittsburgh

Version of record first published: 10 Jul 2012.

To cite this article: Michael J. Ford (2012): A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning, Cognition and Instruction, 30:3, 207-245

To link to this article: http://dx.doi.org/10.1080/07370008.2012.689383