

# Making Classroom Assessment More Accountable to Scientific Reasoning: A Case for Attending to Mechanistic Thinking

ROSEMARY S. RUSS

*School of Education and Social Policy, Northwestern University, 2120 Campus Drive, Evanston, IL 60208, USA*

JANET E. COFFEY

*Department of Curriculum and Instruction, 2226M Benjamin Building, University of Maryland, College Park, MD 20742-1175, USA*

DAVID HAMMER

*Department of Curriculum and Instruction and Department of Physics, 1310 John S. Toll Physics Building, University of Maryland, College Park, MD 20742-4111, USA*

PAUL HUTCHISON

*Department of Education, 1120 Park Street, Grinnell College, Grinnell, IA 50112-1690, USA*

*Received 28 May 2008; revised 26 August 2008, 29 August 2008; accepted 29 August 2008*

*DOI 10.1002/sce.20320*

*Published online in Wiley InterScience (www.interscience.wiley.com).*

*Correspondence to:* Rosemary S. Russ; e-mail: r-russ@northwestern.edu

This work was completed at the University of Maryland, and the ideas expressed in this paper were refined through lively, often heated and contested, discussions among the authors and members of the University of Maryland Physics and Science Education Research Groups.

Contract grant sponsor: National Science Foundation.

Contract grant numbers: ESI-9986846, REC-0440113, and ESI-0227557.

Any opinions, findings, conclusions, or recommendations in this article are those of the authors.

**ABSTRACT:** When teachers or students assess the quality of ideas in science classes, they do so mostly based on *textbook correctness*; ideas are good to the extent they align with or lead to the content as presented in the textbook or curriculum. Such appeals to authority are at odds with the values and practices within the disciplines of science. There has been significant amount of attention to this mismatch in the science education research literature, primarily with respect to experimentation and argumentation as core disciplinary means of assessing ideas. In this article, we call attention to another aspect of scientific reasoning: a focus on causal mechanisms in explaining natural phenomena. We highlight examples and research from the history and philosophy of science to clarify what scientists mean by “mechanism” and to make the case for its centrality. We then present an excerpt from a second-grade class in which a student provides an incorrect mechanistic explanation, and the teacher gives priority to textbook correctness. As the conversation proceeds, the student shifts from mechanistic sensemaking to quoting terminology she does not understand. We argue that attention to mechanism in the classroom would better support student reasoning and better reflect disciplinary epistemology. © 2008 Wiley Periodicals, Inc. *Sci Ed* 1–17, 2008

### INTRODUCTION

Every day, in science classrooms across the country, students express ideas and their ideas are evaluated, by the teacher at least and often by other students. The ideas might come out verbally in class discussions, in writing on homework assignments, or in selections of multiple-choice answers on examinations. The teacher may formally assign the idea a point value, or she may only arrive at an informal, tacit judgment. In classes rich with discussion, the tacit judgments are ubiquitous to the extent that teachers may not even notice them, but they play a significant role in how classes proceed. If a teacher thinks “that’s a good idea,” she responds to it differently than if she thinks “that wasn’t quite what I was looking for.”

Much of this is treated in the education research literature under the topic of *assessment*, and much of that literature focuses on the forms and purposes of assessment (Black & Wiliam, 1998; National Research Council [NRC], 2001a, 2001b). Our primary concern in this paper is on the *substance* of assessment, regardless of the form. Roughly put, we are primarily interested in the criteria teachers, curricula, and—following their lead—students use to decide which ideas are “good” and why.

### Accountability to Curriculum

In many science classes, teachers and students judge the quality of ideas by comparing them to the canon as represented by the curriculum. In other words, making a judgment about whether the idea is right or wrong. Thus they evaluate how well a student’s response aligns with the textbook’s answer, making the textbook or curriculum the authority on accepted science knowledge. We refer this type of judgment as *textbook correctness*, the assessment practices of holding ideas accountable to the factual knowledge in the textbook or curriculum.

Teachers’ and students’ attention remains primarily on textbook correctness for a variety of reasons. For one, it is simpler to compare an idea to the textbook’s answer than to analyze the reasoning that supports that idea. To assess the quality of an idea based on the evidence and arguments that support it can be subtle and difficult; it requires, for example, considering the evidence available to the students. Second, what is tested on external examinations significantly shapes teachers’ classroom assessment practices (Darling Hammond, Aness, & Falk, 1995); teachers naturally pay attention to textbook correctness because that is what they and their students are accountable to in high stakes tests. Perhaps most important,

however, is that to assess the quality of students' reasoning by the methods of science would sometimes mean favoring ideas at odds with the canon, or rejecting ideas within it. This last point, the notion that teachers may need to value ideas that are incorrect when judged against the canon, is both important and—we believe—controversial and is one to which we return throughout this article.

### **Accountability to Science Disciplinary Practice**

Science educators have long recognized that textbook correctness contrasts with the practices of science (e.g., Hodson, 1988; NRC, 2007). Scientists do not believe ideas by authority; they believe ideas because of the evidence and arguments that support them.

Our general purpose in this article is to promote that shift of emphasis and attention *in practice* that science educators recognize *in the abstract*, toward disciplinary practices of assessing the quality of ideas. Especially for early experiences in science, we argue that it is important for teachers to judge the quality of students' ideas in ways that are continuous with practices within science. In fact, we argue that the primary emphasis in elementary science should be on students' developing scientific approaches to reasoning, and to that end teachers ought to assess student ideas primarily by the criteria research scientists use to evaluate the quality of their ideas. This is not to say teachers should ignore textbook correctness; it is to say that that form of accountability should have a lower priority than accountability to disciplinary practices of assessment. We have two reasons for taking this stand.

First and simply, we want students to experience the practices of science. Education reform efforts advocate for the inclusion of inquiry (NRC, 2000, 2007), specifically experimentation (e.g., Kuhn, 1989), argumentation (e.g., Driver, Newton, & Osbourne, 2000), and modeling (e.g., Lehrer & Schauble, 2005), precisely because those activities are claimed to accurately represent professional science practice. Fundamentally, experimentation, argumentation, and modeling are disciplinary means of assessing the quality of ideas, but if textbook correctness is “what counts”—what is valued formally and informally—students get confusing messages about what it means to do science (Coffey, 2003). Thus only when instruction and assessment practices demonstrate to students that scientific reasoning is a productive activity for science learning will students learn that such reasoning is an appropriate activity for them to engage in during science class.

The second reason teachers must explicitly assess ideas based on the evidence, arguments, and models available to students is because of the essential role they play in students' construction and evaluation of their own understanding. The field of science is focused on learning about the world; in a sense, scientists are expert learners. It is by engaging in the practices of scientific inquiry, or nascent versions of those practices, that students will arrive at rich, deep conceptual understanding. In other words, achieving the primary goals of students developing abilities for scientific inquiry is ultimately in the service of their learning the canon of ideas scientists have developed (Bruner, 1960; NRC, 1999). Using textbook correctness as a primary assessment criterion neither reflects the notion that understanding of ideas hinges on quality reasoning nor does it support students' attempts to engage in that reasoning. If teachers care about students truly being able to make sense of correct textbook answers, they must attend to and assess student reasoning, including reasoning that takes students in other, possibly “incorrect,” directions.

### **“Mechanistic”: A Disciplinary Assessment Criterion**

Our specific purpose in this article is to focus on one aspect of what makes for a “good idea” in science, namely that it is *mechanistic*. Historically, the shift from occult and

essentialist ideas to mechanistic reasoning was critical to the scientific revolution that many consider to mark modern science as it is practiced today (Westfall, 1986). Part of what children need to learn is these sorts of *mechanistic* explanations “count” in the practice of science.

In what follows, we recount an example from the history of science regarding the shift to valuing mechanistic reasoning; we do so to make the case for consideration of mechanistic accounts as one criterion on which to evaluate scientific ideas. We do not claim that it is the only criterion used within professional science to judge the quality of ideas. Nor do we claim that it is necessarily the best criterion to use in evaluating reasoning in all circumstances. For example, it may be appropriate to judge a student’s reasoning based on how well it accounts for data or how well it models a particular constraint-based phenomenon. We briefly compare mechanistic explanations to these other criteria as a way to further describe what we mean by “mechanistic” and to explain when it is an appropriate criterion to use for evaluating ideas.

We then shift to an example from a second-grade classroom where students are exploring why an empty juice box collapses when air is sucked out through the straw. We focus in particular on one student’s explanation—which is incorrect—and the conversation that results when the teacher attempts to help her get to a more canonical answer. We argue the student’s productive mechanistic reasoning is overlooked for textbook correctness, perhaps at the expense of the student’s continued productive reasoning. We do not intend for this single case to “prove” our claims about the appropriate role of mechanistic reasoning in the science classroom—nor do we present any line-by-line discourse analysis (for such an analysis, see Russ, Scherr, Hammer, & Mikeska, 2008). Instead, our purpose with this one episode is to lend intuitive plausibility to our claims and to provoke reflection on the possible unintended consequences of various instructional moves in the classroom. The data presented here are illustrative of more general concerns we believe deserve attention in the science education literature.

Finally, we conclude the paper by examining the consequences of focusing attention, and assessment, solely on textbook correctness.

### **ASSESSMENT IN PROFESSIONAL SCIENCE DISCIPLINES**

We begin from the premise that assessment of student ideas in science classrooms, both by the teacher and by the students themselves, ought to reflect how assessment of scientists’ ideas occurs in professional practice. A useful starting point, then, is to articulate more precisely what assessment in the scientific disciplines looks like: By what criteria does the professional scientific community judge which ideas areas acceptable, plausible, or appropriate as explanations for physical phenomenon? We do not attempt a comprehensive answer to that question here; we explore one particular criterion that has been prominent both in historical and contemporary scientific practice.

### **Insight From the History of Science**

One way to make progress in articulating how ideas come to be accepted in science is to consider historical episodes (see Nersessian, 1992, for examples of this approach). The dynamics of such episodes and the character of the theories involved reflect the assessment criteria used by the scientific community at that time and may extend into today. Consider the example of Isaac Newton’s corpuscular theory of light that he formally proposed in 1704. His theory suggests that

Luminous bodies emit corpuscles that are tiny point particles of light with no mass. The properties of light (as observed at that time) can be explained by applying Newton's laws of motion developed for objects with mass to these corpuscles. Specifically, light reflection occurs in the same way balls bounce off walls: the surfaces exert a normal (contact) force in the opposite direction on the moving light particles, changing the direction of their acceleration. Light refracts at the boundary between materials because light particles are accelerated to different speeds in different materials.

What about Newton's theory allowed it to be accepted by his contemporaries as a plausible explanation for light phenomenon? A brief look at scientific practice during that time provides some clues.

Newton presented his theory during the 17th century scientific revolution, a time when the ideas and aesthetics that serve as the foundation for modern science as practiced today were born (Westfall, 1986). During that time, the criteria used to judge the quality of scientific explanations changed dramatically. "Occult" explanations that instilled matter with life and perception were no longer acceptable, and mechanical accounts became the ideal (Westfall, 1986):

The world is a machine, composed of inert bodies, moved by physical necessity, indifferent to the existence of thinking beings . . . all phenomenon of nature are produced by particles of matter in motion. In a mechanical universe shorn of active principles, bodies could act on one another by impact alone. (pp. 33–34)

Shapin (1996) describes how these "mechanical accounts of nature were widely recognized as the goal and the prize" (p. 30). Other historical examples support the suggestion that one of the premises of modern science, as characterized by the scientific revolution, is that appropriate explanations of physical phenomenon should be grounded in mechanistic accounts of nature (Nersessian, 1992; Teeter Dobbs & Jacobs, 1995).

Newton's contemporaries accepted his theory as viable at least partially because it provided a logical, mechanistic picture of the physical world. His theory suggested a mechanism for light phenomenon that was consistent with his experience with the motion of matter in the macroworld, specifically the behavior of massive objects undergoing acceleration due to external forces. Thus "mechanistic" appears to be one criterion that professionals during the scientific revolution began to use to assess the plausibility of explanations.

Even today, when scientists know that Newton's theory is incorrect, his theory and ideas are still valued for making scientific contributions to the study of light that made possible other advances in that area. Darden's description (1998) of four eminent scientists whose ideas were later judged to be incorrect alludes to one reason why correctness should not be the sole driver of the evaluation of ideas. She claims,

Scientific inquiry is an on-going process of error correcting—constructing plausible hypotheses, generating as many plausible rivals as possible, designing new experiments, correcting errors in hypotheses in the face of anomalies.

Science is the process of successively constructing, judging, and refining plausible theories and those theories may at many points along the way be incorrect but still fruitful for making sense of the natural world. Such was the case with Newton's theory of light; his theory was fruitful because it allowed well-reasoned predictions to be made about specific light behavior that could be empirically tested. This insight extends to the classroom.

Judgments of plausibility are important because in disciplined reasoning lies the *potential* for understanding new phenomenon and advancing students' scientific knowledge.

### Insight From Contemporary Philosophy of Science

Insights from contemporary study of sciences show that the scientific community continues to value explanations for phenomenon that are mechanistic<sup>1</sup>: “in many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism” (Machamer, Darden, & Craver, 2000, p. 1).

Other scholars echo the centrality of mechanism to causal explanations (Glennan, 1996; Tabery, 2004). Salmon (1978) suggests that it is only by virtue of being mechanistic that explanations can provide understanding and actually explain. Consistent with the ideals of the scientific revolution, the contemporary scientific community continues to assess whether explanations are acceptable or plausible in part by the extent to which the explanations provide mechanistic accounts of natural phenomenon.

### Unpacking the Criteria of “Mechanistic”

While historical and contemporary accounts of scientific practice suggest the importance of mechanistic explanations, the notion of “mechanism” has been refined and extended since the time of Newton to more accurately capture the work of current scientists (Bechtel & Abrahamsen, 2005; Glennan, 1996; Salmon, 1978; Tabery, 2004; Thagard, 1998). What constitutes an acceptable mechanistic explanation is no longer strictly limited to inert matter in direct contact with other inert matter.

Machamer et al. (2000) suggest more specific criteria that the scientific community might tacitly use when judging whether an explanation is “mechanistic” and thus appropriate for natural phenomena. They describe much of science as the search for mechanisms that underlie stable and reliably produced phenomena. The mechanisms themselves are defined as “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (Machamer et al., 2000, p. 3).

Scientists' goal “in establishing and displaying mechanism is to show how one stage produces the next, and so on” (Machamer, 2004, p. 35). Thus a mechanistic explanation is one that describes this underlying mechanism and each of its sequential stages.

The central components of mechanisms, and thus of mechanistic explanations, are the entities and activities. Activities are the components of mechanisms that produce change—they are the “things that entities do” (Craver, 2002, p. S84). Entities are the things that engage in those activities. The general properties and spatial organization of the entities involved determine the activities that can occur in the mechanism. It is the entities and activities that constitute the mechanism and produce the phenomenon of interest. Thus one possible criterion for whether an explanation is mechanistic is whether it specifies the activities as well as the entities with their relevant properties and organization.

Not only must scientists articulate the entities and activities in a mechanism, they must also describe how each stage of the mechanism progresses to the next stage. Mechanistic explanations do not consist of a series of unconnected, chronological phases; instead each phase necessarily follows from the one before it and necessarily leads to the one after it.

<sup>1</sup>There is much debate in the philosophy of science literature about the purpose and function of explanation in science. What we describe here is one strand from that debate that places primacy on the mechanistic nature of explanation. We do not mean to suggest that this is *the* answer to the debate or that it is the only criteria on which explanations are judged. We only want to point out that it is one aspect of explanation that is valued in professional science.

Darden and Craver (2002) and Darden (2002) call this process of linking each stage to those around it “chaining.” During chaining, scientists “spell out how the previous stage or situation produces this certain result rather than some other” (Machamer, 2004, p. 31). They use knowledge about the entities and activities at one phase to make claims about entities and activities in other phases. Acceptable mechanistic explanations answer—though not explicitly—the following questions, linking together the stages that lead to the observed phenomenon (Darden & Craver, 2002):

- What activities gave rise to entities with these properties?
- What entities are necessary in order for this activity to occur?
- What activities will these particular entities engage in?
- When this activity occurs, what changes will occur in the surrounding entities?

How well a given explanation answers these questions may serve as another criterion on which scientists assess whether explanations are mechanistic. Unless scientists can “chain” to explain how one stage of the mechanism necessarily gives rise to the surrounding stages, they have not provided a plausible mechanistic explanation of the phenomenon.

### **Mechanistic Criterion and Its Relation to Other Criteria**

The mechanistic reasoning we describe is related to several other forms of reasoning discussed elsewhere in the science education literature. In particular, it overlaps substantially with causal and model-based reasoning, both of which we consider to be important for learning in science. There are differences, however, that are important to clarify.

Gopnik and her colleagues (e.g., Gopnik & Sobel, 2000; Gopnik, Sobel, Shulz, & Glymour, 2001) have carried out recognized work on children’s causal reasoning. They designed a “blicket detector” that lights up and plays music whenever “blickets” are brought near it. Their studies have shown that children aged 2–4 can categorize blickets based on their novel “causal power” to set off the detector; that is, even very young children can reason causally to identify which objects (“blickets”) cause the detector to go off.

Mechanistic reasoning is related to this more general notion in that it involves reasoning about the causes underlying physical phenomena. However, it involves more than just reasoning about causality itself—it is more than identifying the “X” that causes “Y” to happen. Mechanistic reasoning also requires that students think about *how* “X” brings about “Y.” In fact, a number of authors in science education have defined mechanistic reasoning precisely this way (e.g., Abrams, Southerland, & Cummins, 2001; Metz, 1991). For example, Koslowski (1996) claims that mechanistic reasoning “explains *the process* by which a cause brings about an effect” (p. 13, emphasis ours). Causal reasoning serves as a starting point for the pursuit of underlying mechanistic explanations, but causal reasoning alone does not define it.

Model-based explanations are also heralded in the science education literature as valuable for the learning and teaching of science. However, there is some variety in the meanings of “models” and “model-based reasoning” across researchers. In their work helping students develop modeling skills, Lehrer and Schauble have described models as analogical representations of various kinds (inscriptions, computational, and so on) that are “extensible, general, and mathematically powerful” (Lehrer & Schauble, 2005, p. 383). In Windschitl, Thompson, and Braaten’s (2008) review of current definitions of modeling in school and professional science, he calls models “testable, modifiable representations of scientific ideas” about the natural world (p. 956). For Stewart, Passmore, and their colleagues models are “a set of ideas that describe a natural process” (Passmore & Stewart, 2002, p. 188);

they are a set of relationships in the natural world that allow students to explain or predict using causal narratives of events (Stewart, Cartier, & Passmore, 2005). Across the literature, model-based explanations generally make use of diverse representations that capture key, underlying features of the phenomenological context, and support causal explanations of that context (Windschitl et al., 2008), although there is not agreement about whether those representations are cognitive or material (see Passmore & Stewart, 2002).

Mechanistic explanations are a particular subset of model-based explanations; not all model-based reasoning centers around understanding the process of how underlying *physical* causes bring about effects. Models may be mathematical, graphical, analogical, physical, or computational (Lehrer & Schauble, 2005), and reasoning about these different models may not require reasoning about the physical system itself. For example, consider the ideal gas law,  $PV = nRT$  (where  $P$  = pressure,  $V$  = volume,  $n$  = number of molecules,  $R$  = a constant,  $T$  = temperature), which is a model of the behavior of an ideal gas. Reasoning with this model is not necessarily mechanistic. A student may treat it as a mathematical constraint, explaining the change in pressure of the gas purely by appealing to its position in the equation ( $P$  is on the same side of the equation as  $V$  so the pressure changes inversely with volume, and so on). Alternatively, a student may reason mechanistically about pressure as caused by the numerous collisions of particles in the gas, explaining the change in pressure of the gas by considering what would affect the frequency and intensity of those collisions (a smaller volume would mean more frequent collisions, and so on). We consider both of these types of model-based reasoning useful for science learning—the first allows the student to make predictions about and describe gas behavior, whereas the second allows him to manipulate or explain that behavior in the physical world and extrapolate to new cases based on an understanding of the mechanism underlying the mathematical model. It is this second, more particular type of model-based reasoning that we distinguish as mechanistic by drawing attention to the nature and features of the reasoning itself rather than to the object (either a material or cognitive representation) on which that reasoning is based.<sup>2</sup>

## ASSESSMENT IN SCIENCE CLASSROOMS

Assessment practices and criteria of professional science disciplines shed light on criteria that teachers and students might use when evaluating student explanations for physical phenomena. In particular, to evaluate explanations in ways that are accountable to science disciplines, educators and students often must be attentive to whether the underlying reasoning is mechanistic. We can also infer that textbook correctness alone does little to judge the fruitfulness of ideas for making progress in understanding new areas of science, since many incorrect ideas laid the foundation for further progress.

To be clear, we recognize that school science and professional science are not one in the same. Therefore, there are times when school assessment must serve some functions where textbook correctness is a reasonable criterion, for practical reasons if for no other. Accountability measures that intend to capture student achievement at a particular point in time, where careful consideration of reasoning could be time intensive and costly, is one such example. Even within everyday classroom assessment, there are times when further exploration or sense making becomes somewhat contingent on an accurate understanding

<sup>2</sup> Windschitl et al.'s (2008) description of models as involving underlying causal mechanisms that incorporate science content aligns most closely with our notion of mechanistic explanations. However, the general description of models within the literature still—at least tacitly—includes other more constraint-based types of reasoning.

of scientific phenomena. However, these occurrences are not the whole of assessment. A primary purpose of assessment is to support student learning (Black & Wiliam, 1998). For this purpose, the more tightly assessment aligns with disciplinary assessment, the more productive it becomes in providing this support for student learning of the discipline.

If we accept that everyday assessment in school science should be accountable to the discipline and that this accountability includes attention to mechanism, we are now prepared to consider the practical implications for classroom instruction. Specifically, what does this mean for what a teacher listens for, recognizes, and responds to? How might classroom discourse directed by the teacher look different when assessment is primarily focused on encouraging mechanistic explanations rather than on textbook correctness? And what are some possible consequences of either focus in terms of student participation and behavior?

In other work (Russ et al., 2008), we explored students' mechanistic reasoning using a discourse analysis framework we developed based on the philosophy of science. Here, we discuss a single example from classroom discourse that we analyze more qualitatively, using the same constructs without the line-by-line coding methodology we employed there. We do not claim that this qualitative case study of this single episode definitively answers our questions—it merely illustrates the possible space of answers and their consequences.

### **A Student Explanation for Phenomenon**

We consider an episode of second-grade students discussing why empty juice boxes collapse when you suck air out through their straws. The conversation took place in a public elementary school in Montgomery County, Maryland, among a science specialist and seven students. In addition to their regular science lessons with their classroom teacher, these students also met several Fridays throughout the year for 50-minute science enrichment sessions.

In this enrichment context, the teacher was free to choose topics based on what he thought would engage the students. The teacher chose to have students discuss why empty juice boxes collapse when they suck the contents—the juice and air—out through their straws. He speculated the students would have productive intuitions about this question based on their everyday experiences. In reflecting after the lesson, the teacher was pleased with the ideas the students generated and pursued in their conversation.

Early in the discussion, after other students have presented several ideas, one student, Erin, gives the following explanation for the juice box collapse:

Erin: I think because since you sucked out the air, it's like, it caves in because there's not any air so it has no, nothing's pushing it in from the inside to make it like [?flat] -

Teacher: Like this?

Erin: - like its normal shape. Yeah.

Teacher: Nothing's inside there so -

Erin: There's not much, not as much is inside so it's, it, there's not mu, as much pushing out so it caves in.

Teacher: Oh. So you mean right now there's air in there pushing out to make it the box shape.

Erin: I think so.

Teacher: And then what happens when I suck? What's -

Erin: You take some of the air out so -

Teacher: - and so why should the sides. (Side comment to another student.) Why should the sides then cave in? I mean, is there anything pushing -

- Erin: No –  
 Teacher: - the outside in?  
 Erin: - there's nothing pushing.  
 Teacher: There's nothing pushing.  
 Erin: So when –  
 Teacher: Nothing pushing where, on the inside or the outside?  
 Erin: Inside.

In explaining the phenomenon, Erin focuses on the role played by the air located inside the juice box. She describes the air inside as actively pushing out on all sides of the box holding them out and flat. When that air is removed from the box (by sucking it out through the straw), there is no longer anything pushing from the inside to hold the box out, so the sides cave in. We call this description of the phenomenon the “inside-pusher” model because of its focus on what happens inside the box.

### Instructional Implications of Possible Classroom Assessments

Using Erin's explanation as a starting point, we now consider our questions regarding assessment—what should the teacher be listening for, recognizing, and responding to in Erin's explanation? Below we discuss what classroom discourse might look like when assessment is primarily focused on encouraging mechanistic explanations, and how that discourse might look different when assessment uses textbook correctness as the primary measure of the quality of student ideas.

**Attention to Textbook Correctness.** For some educators, the primary goal for any given science lesson is for students to learn the textbook material; thus the primary way to assess student ideas is to compare student responses to accepted canonical knowledge as represented in curriculum. Other skills may be important in achieving the goal (e.g., experimentation, argumentation, modeling), but their role is largely one of support. Hammer (1995) describes this stance as

*traditional content-oriented*, because it assesses student contributions with respect to what is traditionally seen as the content of the course. Traditional content-oriented evaluation pertains to the correctness of students' reasoning vis-à-vis an accepted body of knowledge. (p. 403, emphasis his)

From a traditional content-oriented stance, where the main concern is whether students obtain a canonical understanding of air pressure, Erin's explanation is close, but not quite correct. The traditionally accepted explanation for the juice box collapse involves describing the balance and subsequent imbalance of the *two* forces acting on the side of the box. The air inside the box pushes out on the sides, and the air outside the box pushes in on the sides with equal strength. When the air inside is removed there is less air pushing out; the outside force is greater and collapses the box. If textbook correctness is the primary criterion used to assess student ideas, then Erin's explanation is only half right; she lacks a description of the air outside the box. Were Erin to turn in the explanation articulated above at the end of class, such a teacher might give her partial credit since her idea is well reasoned, but would certainly deduct points for the lack of fully correct science knowledge. Although the teacher might recognize some value in her reasonable account of the phenomenon, it is valuable only insofar as it is correct.

For a teacher evaluating student ideas with respect to textbook correctness, a likely initial reaction to an incorrect explanation would be to consider instructional moves that might move the students toward the right answer before the end of the lesson. In this episode, the teacher might respond to Erin by saying “You’re almost there!” and then help her identify the missing conceptual pieces of her explanation. For example, he might ask Erin whether air exists all around the room and if so, what that air does near the box. Alternatively, he might tell her directly that her response needs to account for the air outside of the box as well as inside. Attention to textbook correctness prompts teachers to respond to incorrect student comments with specific conceptual feedback rather than commentary on the appropriateness or plausibility of their idea.

**Attention to Mechanism.** Alternatively, a primary target for student inquiry may consist of learning both how to construct mechanistic explanations and the scientific value in doing so. While textbook correctness may be an overarching goal of instruction, it is not necessarily one that needs to be reached immediately. In fact, focusing exclusively on correctness may contribute to overlooking opportunities to support reasoning that can help students come to better understand correct ideas, thus actually undermining goals of textbook understanding. A teacher who values reasoning might encourage students to spend substantial time grappling with ideas until they make sense based on other experiences with the physical world, even if it means the students leave that class period still not fully understanding canonically correct answers.

From this perspective, a teacher might interpret Erin’s idea as on target because she gives a coherent mechanistic account of the juice box’s collapse. Her explanation includes all of the features of appropriate mechanistic explanations outlined in the above description of the philosophy of science literature. As one measure of its mechanistic character, we observe that her account of the juice box collapsing includes entities (air) organized into particular locations (inside the box) that act in certain ways (pushing out on the sides of the box) to produce the phenomenon. In addition, Erin’s explanation of the juice box is more than a chronological story; she “chains” her story together by describing how one phase necessarily gives rise to the next phase. She answers the question “When this activity occurs, what changes will occur in the surrounding entities?” by explaining that when air inside the box is removed, the box must collapse because that air is what was holding the box flat in the first place. The phenomenon (box collapsing) is the necessary outcome given the entities (air) and activities (pushing out) her explanation described. Erin’s explanation is mechanistic and thus meets an important criterion used by professional scientists when assessing the acceptability of explanations.

For a teacher attending to whether students’ ideas are mechanistic, feedback may help the students recognize the sophistication of the kind of explanations they are giving. The teacher might respond to Erin by saying “Your idea makes a lot of sense!” and explain why, or ask a classmate to explain why. For example, he might ask her whether there are other instances from everyday life that are similar to her idea about air supporting the juice box. If she cannot come up with one on her own, he might offer the example of a pillowcase collapsing when the pillow supporting it is removed as similar to her idea about air supporting the juice box. Then they could talk about the ways in which these examples are similar and different. Another possible response would be to take up her reasoning: “How does air inside the box ‘push’?” he may ask her. After getting a number of student ideas on the table, he could encourage students to compare and contrast the differing explanations, and then give them a chance to revise their stories.

The range of possible productive responses show that teacher feedback of this nature is different than targeted “correctness” feedback. For one, feedback of this kind sends Erin and her classmates the tacit epistemological message that drawing on personal experiences with mechanisms is an appropriate thing to do in science class. Such a teacher values and encourages mechanistic explanations in and of themselves as productive means for making sense of physical phenomenon regardless of whether the explanation is fully correct. Assessing ideas based on their mechanistic character prompts teachers to respond to incorrect student comments in the same way he responds to correct ones, by providing feedback about whether they are constructing explanations appropriate for science and fruitful for further progress in understanding new knowledge. Teachers with this perspective might only provide conceptual feedback *after* (if at all) the students have developed a stable mechanistic approach to science. Overall, they would be pleased with the reasoning that Erin displays.

### Pitfalls of Classroom Assessment Centered on Textbook Correctness

The analysis above shows that a teacher’s aims can lead to dramatically different interpretations of Erin’s explanation. For a teacher assessing her explanation using the criterion of textbook correctness, Erin has *almost* reached the goal but needs to be pushed a little further conceptually. In contrast, for a teacher assessing her explanation based on its mechanistic character, Erin has *already* reached a major goal and only needs to be shown that her explanation is appropriate and fruitful for science learning. These different interpretations influence how teachers respond to students, which in turn can influence how students engage in inquiry.

We now look at the consequences of various assessment practices. We use actual discourse data from Erin’s classroom to describe some *possible* outcomes of assessment on student behavior. In particular, we suggest that significant negative consequences for student participation and epistemology may ensue if classroom assessment privileges textbook correctness over mechanism as the primary measure of quality.

In the episode with Erin, the teacher’s primary focus on textbook correctness led him to repeatedly push Erin to give the correct answer even when she had developed a highly mechanistic incorrect explanation that is plausible and appeared to make sense to her. This instructional move has an immediate impact on how she engaged in inquiry in the moment and may also send an inaccurate message about productive approaches to learning science.

In the episode presented above, Erin responded to the teacher’s question with a well-articulated “inside-pusher” explanation that is mechanistic. Erin continued to describe her mechanistic account using ideas that appeared to make sense to her even after the teacher hinted at the correct answer by asking “Is there anything pushing the outside in?” When the teacher then asked her to explain more fully, we see a startling transition in her approach to the task:

- Teacher: So why, if there’s nothing pushing on the inside, why should the outside, why should the box’s sides cave in? [3 second pause]  
(Unrelated comment to another student)
- Erin: ‘Cause uh, there’s not as much air in [this/it?].
- Teacher: Okay, so there’s less air in the inside that way.
- Erin: Yeah.
- Teacher: But I don’t understand why that makes the sides have to go in.
- Erin: [5 second pause] Maybe its pressure I don’t know.
- Teacher: What’s that? Pressure?

- Erin: It's something that's hard to explain. Um. [4 second pause] It's something that's [6 second pause] like, it's hard to explain.
- Teacher: Okay. Let's try, as a group and individually.

What is most notable about this episode is the change in the nature of Erin's explanation from her previous mechanistic one. The first time the teacher asked for clarification, Erin quickly responded with a meaningful restatement of her mechanism. Erin read his question as a request for a causal story and responded accordingly; she was still trying to make sense of the situation. However, the second time he asked, Erin spent 5 seconds considering his question before changing the content and type of response. Instead of continuing with her mechanistic idea, she responded to the teacher's repeated question by adding the term "pressure" to her explanation in what seems to be an attempt to satisfy him. Her halting response and long pause to his question "What's that? Pressure?" suggests that she was using the term without understanding what it means.

The task at hand has shifted, namely from explaining what caused the juice box to collapse to explaining the term pressure. We posit that what drove this shift was the teacher's push for Erin to reach the correct explanation. This instructional push for textbook correctness caused Erin to shift from constructing a coherent mechanistic explanation with entities and activities that made sense to her (air and pushing) to invoking technical vocabulary she does not appear to understand (pressure). Erin does not link the idea of pressure into her previous mechanistic story; she does not chain together "pressure" with the elements of her "inside-pusher" model. Her more productive explanation strategies are suppressed in the moment by a desire to reach textbook correctness.<sup>3</sup>

Following Erin's introduction of pressure, the teacher turned the conversation away from making sense of how pressure fits into her previous mechanistic story and toward defining the term pressure "as a group and individually." In doing so, he may tacitly (and inadvertently) confirm Erin's suspicion that correct vocabulary was what he wanted. We do not claim that the teacher was explicitly looking for scientific vocabulary; he may be searching for a correct answer, which in Erin's case manifests itself as terminology. Instead, we suggest that Erin (and the other students) may interpret his question and response in that way. As her response indicates, the teacher's response to her pressure idea may send Erin the epistemological message that scientific inquiry is about producing correct, "scientific-sounding" answers rather than about constructing causal mechanistic accounts for phenomenon. Erin may learn that textbook correctness is what counts, and that mechanistic ideas are neither appropriate nor valuable for science. As a result, she may be less likely to engage in constructing such explanations when she is asked to learn scientific concepts in the future.

Interventions such as this one that are driven primarily by a desire for students to learn the right answer might ultimately damage teachers' broader goals of conceptual understanding. As Hammer (1995) describes,

Instructors may too quickly undermine their own objectives for students understanding by insisting too quickly on correctness. . . Students may learn to produce correct statements without developing understanding. (p. 427)

If students are pushed too quickly toward a correct answer that does not make sense to them over an incorrect answer that does make sense to them, they may accept that correct

<sup>3</sup> It is possible that the teacher is only pushing Erin in this way because he thinks her sophisticated reasoning indicates she is prepared to go farther with her explanation. However, the result remains that his questions pushed Erin out of a (fairly stable) mechanistic reasoning mode.

statement without attempting to understand it. In the air pressure discussion, the teacher's focus on immediately achieving textbook correctness forces Erin out of a productive explanatory mode that would help her make sense of that correct knowledge. In addition to suppressing mechanistic explanation in this episode, his attention to correct vocabulary may also discourage Erin from constructing explanations in future inquiry that would help her gain a meaningful understanding of science knowledge. In general, if teachers allow the desire for textbook correctness to overcome the desire for quality mechanistic explanations in every moment they may lose access to the very explanations that would support an understanding of correct canonical concepts.

## REVISITING ASSESSMENT

In the analysis above we argue that teachers' assessment of reasoning is critical for maintaining accountability to professional science. Examples from history of science demonstrate that it is the reasoning supporting new ideas that allows scientists to decide whether those ideas will be fruitful and productive for further understanding the physical world. In particular, mechanistic reasoning, among other things, has been shown to be valuable for the advancement of science knowledge.

We argue that the same should be true in science classrooms. Teachers and students should primarily evaluate new ideas based on the character of the underlying reasoning. But reasoning-based assessments are not only important from the philosophical view that science class should mirror science practice; such assessments are also important *for students* both practically and epistemologically. Practically, sound reasoning supports meaningful understanding of correct ideas; students will not have opportunities to develop and refine that reasoning unless it is assessed and valued. From the epistemological view, unless assessment indicates to students that reasoning is important independent of the answer it produces, students may tacitly learn that such reasoning is inappropriate or unproductive for science learning.

In the abstract our claims seem easy to accept, but we use the case with Erin to illustrate and push the point to the fore. If we accept that reasoning is valuable—for accountability to the discipline, for supporting meaningful correct understanding, and for conveying an accurate message about the nature of science—then productive reasoning must be pursued *even if it leads to incorrect answers*. Erin is just such a case where productive mechanistic reasoning leads her to an incorrect answer for the juice box question. In that moment, her teacher makes the tacit decision to pursue the target of textbook correctness. Our point with this example is to show how quickly a teacher can slip into a mode of assessment centered on textbook correctness and in turn how quickly that mode can have negative consequences. In particular, Erin quickly stops reasoning in a way that might support her understanding of the correct ideas and she likely also tacitly learn something about how to engage in these discussions in the future.

Our concern is this: If we allow our desire for textbook correctness to dictate our assessments in every moment, we risk undermining our overarching goal of developing meaningful understanding of science and scientific knowledge by suppressing the reasoning that leads to it. This is not to say that correctness never matters; there are times in the classroom when we want or need students to have correct answers, especially when those answers are prerequisites to further exploration of content. However, we suggest that while textbook correctness is certainly a reasonable overarching goal for science learning, it may often be an inappropriate *immediate* goal in the classroom. To adeptly negotiate this tension, textbook correctness cannot serve as the only, or even the primary, criterion by which we judge each individual contribution. Rather overarching goals for accurate

textbook understanding ought to grow out of immediate attention to reasoning practices that will support that goal.

While the primary focus of this paper has concerned what teachers pay attention to and support with respect to assessment, implicit are important consequences for students' emergent ideas of science. For students, assessment helps identify what is "important" in their school subject areas (Lampert, 2001). In this way, it marks and defines the discipline. If our immediate or sole focus is correctness of knowledge, that is what students will come to understand science to be: Science is about producing correct answers using the correct terminology. If we broaden our focus to include the reasoning and practices that constitute the discipline because they are productive for constructing that knowledge, then what is conveyed is a very different idea of what science is, one that is much more grounded in the discipline.

Many students come to our classrooms not knowing, or believing, that mechanistic reasoning is an important part of science and scientific inquiry. Like what Erin in the example above did when she shared her "inside-pusher" explanation for the collapse of her juice box, they often construct mechanistic stories even while they may not be aware that engaging in this type of reasoning is something that is valued. As a second grader, it may even be safe to assume that Erin did not recognize that what she was doing was an appropriate way to construct meaningful scientific knowledge. And, likely, she would be unable to identify what it was she was doing. Therefore, teachers' responses become critical: Erin, and students everywhere, begin to construct their ideas about what science is in large part through what occurs in science classrooms.

It is impossible to say from the classroom transcript how this exchange will be consequential for Erin's and her classmates' ideas about science or on their participation in further science courses. What we do see in this short segment, however, is a significant shift in how Erin participated in the activity—both in the nature of her reasoning and in her affect. She went from confidently offering a reasonable story about how the box collapsed to hesitantly and uncomfortably searching for a term she did not understand to satisfy the teacher's continued push for explanation. This single episode may not have a long-standing influence on Erin and her classmates; however, an accumulation of school experiences will likely shape her ideas of what it means to do science. In focusing on textbook correctness, we perpetuate students' unawareness of this building block of meaningful conceptual understanding; and, importantly, we send a strong epistemological message that it is less important to reason mechanistically and make sense of phenomenon than it is to get the right answer.

To achieve larger goals of conceptual understanding in science, we need to help students establish and maintain a productive stance toward generating science knowledge. Children have a variety of stances they can take toward learning (Hammer & Elby, 2002). Thus an important part of science learning must include helping students understand what stances are valuable for learning science ideas, which involves helping them understand what kinds of things scientists do when they consider and propose ideas. This includes helping them understand what kinds of explanations count as scientific and what criteria are used to evaluate these ideas. Within the science education community, much emphasis has been placed on helping students understand that new ideas must align well with empirical observations. However, if we are accountable to the discipline we cannot stop with that criterion alone. We must also help students develop the understanding that evaluation of ideas in science also requires judgments of reasoning as plausible, sensible, and mechanistic.

We suggest that a firm understanding of assessment criteria used in science disciplines could help students engage more productively in learning science. They will be better able to construct scientific explanations, anticipate arguments, evaluate ideas and claims, and

engage in sound and productive argumentation. In other words, they will be able to engage in doing science. They will also be better prepared to learn canonically correct content, as they will be able to better make sense of new information.

This discussion ties back to the question of what Erin's teacher should do when he hears her explanation. Teachers need to recognize and support the beginnings of sophisticated science. As we discussed above, this involves the teacher attending to student ideas as a scientist would attend to the ideas of a colleague. An important aspect of this support is making meaningful assessment criteria—that which is accountable to the discipline—part of the goals for students' science learning. When students learn to engage in disciplinary assessment practices, they are learning science. And helping them learn science in ways that are authentic to the discipline will allow us to make more headway toward meaningful, conceptual understanding.

## REFERENCES

- Abrams, E., Southerland, S., & Cummins, C. (2001). The how's and why's of biological change: How learners neglect physical mechanisms in their search for meaning. *International Journal of Science Education*, 23(12), 1271–1281.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421–441.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.
- Coffey, J. E. (2003). Involving students in assessment. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 75–88). Arlington, VA: National Science Teachers Association Press.
- Craver, C. F. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science (Supplement)*, 69, S83–S97.
- Darden, L. (1998). The nature of scientific inquiry. Retrieved October 31, 2005 from <http://www.philosophy.umd.edu/Faculty/LDarden/sciinq/index.html>
- Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science (Supplement)*, 69, S354–S365.
- Darden, L., & Craver, C. F. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 33, 1–28.
- Darling Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teachers College Press.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287–312.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49–71.
- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, 71(5), 1205–1222.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5), 620–629.
- Hammer, D. (1995). Student inquiry in a physics class discussion. *Cognition and Instruction*, 13, 401–430.
- Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190). Mahwah, NJ: Erlbaum.
- Hodson, D. (1988). Toward a philosophically valid science curriculum. *Science Education*, 72(1), 19–40.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674–689.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.
- Lehrer, R., & Schauble, L. (2005). Cultivating model-based reasoning in science education. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–387). Cambridge, MA: Cambridge University Press.
- Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science*, 18(1), 27–39.

- Machamer, P., Darden, D., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Metz, K. E. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching*, 28(9), 785–797.
- National Research Council (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. Committee on Developments in the Science of Learning, J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Commission on Behavioral and Social Sciences and Education.
- National Research Council (2000). *Inquiry and the National Science Education Standards*. Washington, DC: National Academy Press. Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry. Center for Science, Mathematics, and Engineering Education.
- National Research Council. (2001a). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press. Committee on the Foundations of Assessment. J. Pelligrino, N. Chudowsky, & R. Glaser (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education.
- National Research Council. (2001b). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academy Press. Committee on Classroom Assessment and the National Science Education Standards. J. Myron Atkin, P. Black, & J. Coffey (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press. Committee on Science Learning, Kindergarten Through Eighth Grade. R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education.
- Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science* (pp. 3–45). Minneapolis: University of Minnesota Press.
- Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching*, 39(3), 185–204.
- Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92, 499–525.
- Salmon, W. C. (1978). Why ask, "Why?"? An inquiry concerning scientific explanation. *Proceedings of the 52nd Annual Meeting of the American Philosophical Association*, 51, 683–705.
- Shapin, S. (1996). *The scientific revolution*. Chicago: The University of Chicago Press.
- Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In National Research Council (Ed.), *How students learn: History, mathematics, and science in the classroom* (pp. 515–566). Washington, DC: National Academies Press.
- Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science*, 71, 1–15.
- Teeter Dobbs, B. J. & Jacob, M. C. (1995). *Newton and the culture of Newtonianism*. Atlantic Highlands, NJ: Humanities Press International.
- Thagard, P. (1998). Explaining disease: Correlations, causes, and mechanisms. *Minds and Machines*, 8, 61–78.
- Westfall, R. S. (1986). *The construction of modern science: Mechanisms and mechanics*. New York: Cambridge University Press.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967.