

The Missing Disciplinary Substance of Formative Assessment

Janet E. Coffey,¹ David Hammer,² Daniel M. Levin,¹ and Terrance Grant³

¹*College of Education, University of Maryland, College Park, Maryland*

²*Tufts University, Medford, Massachusetts*

³*St. Paul School, Baltimore, Maryland*

Received 17 April 2011; Accepted 5 August 2011

Abstract: We raise concerns about the current state of research and development in formative assessment, specifically to argue that in its concentration on *strategies for the teacher*, the literature overlooks the *disciplinary substance* of what teachers and students assess. Our argument requires analysis of specific instances in the literature, and so we have selected four prominent publications for consideration as examples. These, we show, pay little attention to the student reasoning they depict, presume traditional notions of “content” as correct information, and treat assessment as distinct from other activities of learning and teaching, even when they claim the contrary. We then offer an alternative image of formative assessment centered on attention to disciplinary substance, which we illustrate with an example from a high school biology class. Assessment, we contend, should be understood and presented as genuine engagement with ideas, continuous with the disciplinary practices science teaching should be working to cultivate. © 2011 Wiley Periodicals, Inc. *J Res Sci Teach* 48: 1109–1136, 2011

Keywords: formative assessment; teaching; science education

Introduction

Our first purpose in this article is to argue that the literature on formative assessment largely overlooks the disciplinary substance of what teachers—and by emulation their students—should be assessing. Part of this claim concerns the depiction of assessment and part concerns what science education should consider as substance. Compelling the case in both respects, we believe, requires examination of specific examples, and so rather than try for a comprehensive review of the assessment literature as a whole, we have selected four publications by prominent researchers in assessment and science education. We quote extended excerpts from each, including transcript of classroom discussion and what the article or book has to say about it; these excerpts are the evidence we use to develop our case.

We open with one from *Assessment for Learning: Putting it into Practice* (Black, Harrison, Lee, Marshall, & Wiliam, 2003), a highly cited publication by two of the best-known authors on formative assessment, Paul Black and Dylan Wiliam. It is a discussion from sixth-grade science class, which the book offers as a focal example of effective formative assessment.

In Black et al.’s (2003) account, the teacher had presented two geranium plants, one thriving and one not, and asked the students to consider why they might have grown

Contract grant sponsor: National Science Foundation; Contract grant numbers: ESI 0455711. DRL 0732233.

Correspondence to: J.E. Coffey; E-mail: coffey.janet@gmail.com

DOI 10.1002/tea.20440

Published online 7 September 2011 in Wiley Online Library (wileyonlinelibrary.com).

differently. He told the students that both plants started out the same, but that they “may have been growing in different places,” and that “it’s got something to do with the way plants feed” (Black et al., 2003, p. 37). For about 4 minutes, the students spoke in pairs, and then he called them together for a full class discussion, which the book presents as follows:

Teacher: Okay. Ideas?

About half the class put up their hands. Teacher waits for 3 seconds. A few more hands go up.

Teacher: Monica—your group? Pair?

Monica: That one’s grown bigger because it was on the window [Pointing].

Teacher: On the window? Mmm. What do you think Jamie?

Jamie: We thought that. . .

Teacher: You thought. . .?

Jamie: That the big ‘un had eaten up more light.

Teacher: I think I know what Monica and Jamie are getting at, but can anyone put the ideas together? Window–Light–Plants?

Again about half the class put up their hands. The teacher chooses a child who has not put up his hand.

Teacher: Richard.

Richard: Err yes. We thought, me and Dean, that it had grown bigger because it was getting more food.

Some students stretch their hand up higher. The teacher points to Susan and nods.

Susan: No it grows where there’s a lot of light and that’s near the window.

Teacher: Mmmm. Richard and Dean think the plant’s getting more food. Susan. . . and Stacey as well? Yes. Susan thinks it’s because this plant is getting more light. What do others think? Tariq.

Tariq: It’s the light cos its photosynthesis. Plants feed by photosynthesis.

The teacher writes photosynthesis on the board.

Teacher: Who else has heard this word before?

The teacher points to the board.

Almost all hands go up.

Teacher: Okay. Well, can anyone put Plant, Light, Window and Photosynthesis together and tell me why these two plants have grown differently?

The teacher waits 12 seconds. Ten hands went up immediately he stopped speaking. Five more go up in the pause.

Teacher: Okay, Carolyn?

Carolyn: The plant. . . the big plant has been getting more light by the window and cos plants make their own food by photosynthesis, it’s. . .

Jaime: Bigger.

Teacher: Thanks Jamie. What do others think about Carolyn's idea?

Many students nod (Black et al., 2003, pp. 38–39).

We stop here; in the book the case continues for a bit longer: The teacher reiterated Carolyn's idea and affirmed it as correct. He then turned to Richard's and Dean's idea, which Dean said was "wrong" but Richard said they had meant to say the same thing. The teacher asked Richard to say the "idea again but use the word photosynthesis," and he accepted Richard's answer as "not bad" (Black et al., 2003, p. 39).

Black et al. (2003) discuss how these data are evidence of teacher's progress in formative assessment as a result of professional development:

This extract shows a marked difference in the way that the teacher approaches questioning. He is no longer seeking terms and descriptions but rather trying to explore students' understanding. He creates opportunities for the students to exchange ideas, articulate their thoughts and to fashion answers in a supportive environment. Wait time is greatly extended and this encourages more students to participate and think of answers. The students' answers are longer and contain indications of their conceptual understanding rather than their knowledge of names and terms. (p. 39)

We do not doubt that the teacher made progress. As the authors' analysis indicates, he was now eliciting students' thinking, and he provided wait time (Rowe, 1974) before he sought responses. He asked students to consider and build on each other's ideas; he held back on explicit evaluation, disrupting the common pattern of classroom talk (Mehan, 1979). Moreover, what he heard informed his teaching.

At the same time, the example is striking for how it depicts the teacher's attention and responsiveness to student thinking. Monica and Jamie said that the plant was "growing bigger because it was on the window," that it "had eaten up more light," to which the teacher responded that he thought he knew what they were getting at, and "Can anyone put the ideas together?" The problem, however, is that we and the teacher and the students have do not have enough information to understand what those ideas were. Did Jamie think that light is food for plants the way bread is food for people? Genuinely to consider that idea—that is genuinely to *assess* it—would mean raising questions about food and light and eating: Can something be food that isn't a *thing*? Or is light a thing? What about water and soil: Could they be food for the plant? Can a living thing "eat" just by sitting there? And so on. But such questions never come up.

Another possibility is that Monica and Jamie were not focused entirely on *plants*. Perhaps they were thinking about the *teacher* and the hints he had given that the answer involves food and the plant's location. Again, we and the teacher and the students would need to hear more, but if that was what they were doing, it would explain why questions about food and light and eating never came up: The students were focused on getting the teacher's answer, rather than on how that answer fits with the rest of their knowledge and experience.

There is no evidence the teacher noticed, after Richard said "the [bigger] plant was getting more food," that Susan said "No it grows where there's a lot of light," (p. 38) evidently distinguishing *food* from *light*. He noticed and responded when Tariq said "photosynthesis," which was the only thing he wrote on the board during this excerpt. If he genuinely had been "trying to explore student understanding," as the authors contended, he would have spent time looking into and unpacking their ideas. Rather, the evidence suggests, he was more focused on finding out what the students already knew of the target information.

We discuss this excerpt further below, along with three others from other prominent publications, to challenge the current state of research and development in formative assessment, our first purpose in this article. Assessment deals, fundamentally, with teachers' and learners' awareness of ideas and reasoning, and strategies for formative assessment aim at helping teachers attend to what their students are thinking during the course of instruction. However, perhaps out of a desire to have the widest relevance, authors have focused on strategies that cut across topics and disciplines, such as wait time or "stop lighting" or questioning, without closely examining the ideas and reasoning they reveal. By not delving into the specific substance of student thinking, the literature—and, subsequently, practice—misses and may undermine its fundamental objective.

Our second purpose of this article is to offer an alternative image of formative assessment centered on *attention to disciplinary substance*, which we illustrate with data collected from the fourth author's 9th grade biology classroom during a year he devoted to reforming his practice. Although it is difficult to put a name to specific strategies Terry was using, we show how assessment came to inhere in his interactions with students: He was attending and responding to their ideas and reasoning, as a seamless part of teaching. The heart of the matter of formative assessment, we argue, should be seen as that attention, to what and how students' are thinking and participating. Strategies should be in the service of that attention.

The remainder of this article is organized as follows: In the following section, we present evidence that, in consequential ways, literature in formative assessment has neglected disciplinary substance, and we draw from research on learning to argue that this neglect has been problematic. We do not intend a comprehensive review; that would not afford examining examples at the necessary level of detail. Rather, we focus on four examples, all of which depict formative assessment occurring within classroom discussions that elicit student thinking.

We then offer an alternative view, of *assessment as attention to substance*, citing research on teaching (Ball, 1993; Hammer, 1997; Sherin, Jacobs, & Philipp, 2011) that, we argue, is relevant to the topic of formative assessment, although the authors have not spoke of it in such terms. We illustrate this view with an example from Terry Grant's high school biology teaching. We close the article with a discussion of implications for research and professional development.

The Neglect of Disciplinary Substance in Research on Formative Assessment

Much of the emphasis in formative assessment has been in its function and timing, as assessment that occurs *within* learning activities rather than subsequent to them. It provides information the teacher can use to make judgments during a class, for example, or day-to-day in planning lessons. (Black & Wiliam, 1998a; National Research Council, 2001a, 2001b; Ramaprasad, 1983; Shepard, 2000). Formative assessments are often used synonymously with benchmark or interim assessment and in reference to test items (Bennett, 2011; Popham, 2006). Current work in formative assessment has focused on expanding those images to include conversations, our central interest in this article, such as the one from Black et al. (2003) we presented above. In these accounts, as Bennett (2011) summarized, "formative assessment is not a test but a process" (Popham, 2008, p. 6) that "produces not so much a score as a qualitative insight into student understanding" (Bennett, 2011) that can inform instructional decision-making (Shepard, 2000).

One reason for the increased emphases on formative assessment in the literature is the compelling research showing that effective use of formative assessment can have a significant influence on student achievement (Black et al., 2003; Black & Wiliam, 1998a; White &

Frederiksen, 1998). The range of classroom activities captured in these studies is broad, from questioning and feedback practices to journaling and reflection. Some evidence cites large effect sizes (as high as 0.4–0.7 according to Black and Wiliam, 1998b), although not all of the research points to positive effects (Bennett, 2011; Kluger & DeNisi, 1996; Shavelson et al., 2008; Torrance & Pryor, 1998). This body of work on formative assessment has enriched our understandings of how assessment can support student learning.

One theme of these findings is that the nature and quality of the questions teachers pose matter for the nature and quality of the student thinking they reveal and promote (Black & Wiliam, 1998a). Substantial research supports the value of open-ended questions, such as in the excerpt from Black and Wiliam (above, that require and afford more than single word responses). Such questions elicit more information from students (Nystrand, Wu, Gamoran, Zeisler, & Long, 2003), which provides teachers with more data and sparks deeper student thinking (Jos, 1985). Teacher questioning can change discourse patterns of the classrooms, helping dialogue move away from more typical initiation, response and evaluation (Mehan, 1979) to a more free flowing exchange of ideas (Minstrell & van Zee, 2000; Minstrell & van Zee, 2003). Furthermore, different types of questions elicit different types of knowledge available for assessment (NRC, 2001b; Shavelson & Ruiz-Primo, 1999). And the importance of wait time (Rowe, 1974) is now a staple of discussions about best practices.

A second theme of findings in research on formative assessment is that the nature and quality of the feedback teachers give students matters as well, as a central means by which insights from formative assessment can support student learning (Black & Wiliam, 1998a). A variety of studies have focused on how teachers use the insights from formative assessment to give students feedback about their progress, and the evidence shows some feedback to be more effective than other (Black & Wiliam, 1998a; Butler, 1987; Butler & Neuman, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Studies point to the benefits of teacher feedback that is descriptive (specific to task) (Tunstall & Gipps, 1996), offering guidance for improvements (Butler & Neuman, 1995), focused on specific task (Ames, 1992; Dweck, 1986), and close in proximity to when the work was completed (Erickson, 2007).

This has been important work, and it has changed and is continuing to change practices of assessment in schools. The concerns we raise here have to do with how accounts in the literature treat disciplinary substance, as we began to articulate with respect to the example above. Black et al.'s (2003) discussion of the classroom data gave little consideration to the disciplinary substance of the students' ideas and reasoning. (What did the student mean, saying the plant had "eaten up more light"?) Of course, it would be difficult to give student thinking close consideration without also entering into discussion about its disciplinary value, but that is a further problematic aspect of the account. Research on assessment strategies also says little about the substance of the objectives. (What should this instruction be trying to accomplish? What are the practices and understanding that constitute science here?)

Absent critical attention to the disciplinary substance of either student thinking or instructional objectives, the literature tacitly presumes traditional notions of "content" as a body of correct information. Again, this was evident in the opening example, where it is evident the teacher's primary objective was that students learn that plants use light to feed in a process called *photosynthesis*. Formative assessment meant comparing student contributions to that target, and what was significant for Black et al. (2003) in their analysis was that the teacher did this comparison during instruction. With content as a body of information, strategies of assessment can be generic and separable from subject matter, to determine *do they have the information*, whatever that information might be.

As that example illustrates, our concerns arise primary in the instance, in the specific depictions of learning and instruction researchers present to illustrate formative assessment. We recognize that to single out cases for attention begs the question of whether they are representative of the literature as a whole, but it would not be possible for us to provide evidence at the necessary level of detail from more than a few sample publications. We also realize that it may seem indecorous of us to criticize the work of particular individuals. We certainly do not intend anything *ad hominem*. We intend this article in the spirit of open scholarly debate, which is as important within science education research as it is within science, and we would certainly welcome rebuttal.

For these reasons, we have taken care to choose prominent publications by established, influential researchers. We anticipate that readers of this *Journal* will be familiar with the names to follow, but for those who are new to the literature we provide evidence of the influence of these articles from records of their citation in the literature. Black et al. (2003), from which we drew the opening example, has been cited 599 times as of the submission date of this article.¹ A shorter pamphlet version produced by King's College, London (Black, Harrison, Lee, Marshall, & Wiliam, 2004a), has been cited an additional 545 times.² We draw our second example from Morrison & Lederman (2003), work more centrally situated within science education than the others, on conversational assessment of students' "preconceptions"; it has been cited 48 times. The third example is from Bell and Cowie (2001a), which has been cited 124 times; a journal length article of the same research (Bell & Cowie, 2001b) was cited an addition 91 times.³ Finally, we take an example from a special issue edited by Shavelson of *Applied Measurement in Education*, focused on formative assessment, which has been cited 57 times.

Having introduced the article with the first excerpt, we organize our critique around the three further examples to support three interrelated claims regarding the depiction of assessment through conversation in science class:

1. There is little discussion about the substance of student thinking.
2. There is a tacit presumption of "content" as a body of correct information, centered on terminology and selected in advance as lesson objectives.
3. Assessment is discussed in terms of particular strategies, techniques, and procedures, distinct from other teaching and learning activities.

In each case, we support our claims primarily by reexamining the data and analyses the authors have included in their published work. Again, we have chosen for this work to focus on conversational (or "informal" or "on-the-fly") formative assessment (Shavelson, 2003; Shavelson, 2008), in which there is an immediate feedback loop within classroom interactions. Our argument would apply as or more strongly to examples of more formal, instrument-based formative assessment, but in the interest of space we save that case for another article.

Following our discussion of these instances, we review arguments from research on learning to argue that the neglect of substance in the literature is problematic. We then turn to research on teaching for images of conversational assessment that attends and responds to the disciplinary substance of student thinking, and in the final section of the paper we offer an example from our own work.

There Is Little Discussion About the Substance of Student Thinking

We have given one example, from Black et al. (2003), from a middle-school science class. Here is a second, from a high school science classroom, as part of a study on how

Journal of Research in Science Teaching

science teachers diagnose “preconceptions” (Morrison & Lederman, 2003). The authors represented as an example of a teacher attempting to “find out students’ understanding”:

Teacher: You all know what a sunspot is?

Student 1: I don’t know what it is.

Teacher: Okay, who does know what it is so they can answer? The rest of you must fit in that category right?... Mr. Evans, what’s a sunspot?

Student 2: Ahhhh, ummmm

Teacher: You are asking me to ask my questions and answer my questions? No way.

Student 3: A burst of energy.

Teacher: A burst of energy. From what? (pause)

Student 3: The sun.

Teacher: The sun. Why?

Student 3: Because it has gas explosions.

Teacher: Oh... Nope (pause)

Student 4: Different gases, a bunch of different gases burn.

Teacher: What are some of these different gases?

Student 4: Helium.

Teacher: Helium burns real well doesn’t it? Pretty soon you are going to think of the sun as a gas ball and it’s on fire and we get the heat off that fire. Not likely!... Stan?

Student 5: I have two... One is that it is a stretch of energy in the sense that the radioactive waves are coming together and it’s like exploding or it is a matter of fluctuations of energy, the sun has so much energy it is creating we’re gonna have fluctuations.

Teacher: Neither one comes close. Sunspots. Does that imply singular? (pause)

Student 6: Particle flares?

Teacher: Particle flares, solar flares. Sunspots occur in pairs, any hint now?

Student 7: Is it polar?

Teacher: The flares are polar, very definitely. One is a different pole than the other one (Morrison & Lederman, 2003, p. 857).

The researchers describe this example as demonstrating that this teacher (unlike the others in their study) “attempts to find out students’ understanding of the concepts being discussed” (p. 857):

Bill, on the other hand, often kept posing and probing with his questions until students started to provide ideas. He said he wanted to hear what all students had to say, right or wrong. Bill expected the students to express their ideas about the topics discussed; he berated them if they provided no input, saying he would rather have wrong ideas than none at all. (p. 857)

While Bill, the teacher, asked students to express their answers, the evidence does not support the claim that he was doing so to “find out students’ understanding.” His quick rejections of the ideas from Students 3, 4, and 5 confute the notion that he was interested in all ideas, “right or wrong.” Rather, they indicate that he was measuring their contributions against what he took to be the correct answer. He did not focus on the disciplinary substance of the ideas for their own value; nor did he give evidence or reasoning to support rejecting them. The researchers’ analysis treated this as unproblematic, and, in fact, presented it as good pedagogical practice.

Meanwhile, there is validity to what students were beginning to say. The idea of a sunspot as a “burst of energy” arising from “gas explosions” is entirely sensible, connected to the students’ understanding of the sun as a “gas ball” that is “on fire.”⁴ The idea of “fluctuations” must have at least some validity; it seems likely the student was thinking of some element of randomness to the appearance of solar flares.

The example is also problematic with respect to the disciplinary objectives. The teacher took the correct answer to involve magnetism. The excerpt ends there in the paper, so we do not know what followed, but we do know that the researchers considered this an adequate segment. The problem is that a disciplinary understanding would need much more than the matter-of-fact information that sunspots result from magnetic fields: What evidence and reasoning supports this conclusion? And, certainly, what evidence and reasoning could refute the students’ ideas? The article does not provide any discussion about the disciplinary understanding that the teacher is hoping students to achieve, which brings us to our second claim.

There Is a Tacit Presumption of “Content” as a Body of Correct Information, Centered on Terminology and Selected in Advance as Lesson Objectives

Absent discussion about the nature of the disciplinary objectives, researchers tacitly support traditional views of content as information students should retain. In the opening example, the information was that plants use light to feed in a process called *photosynthesis*; there was no genuine consideration of other ideas. In the second, the target information was that *sunspots* arise from *magnetic fields*. The teacher assessed student contributions for their alignment with that; ideas that did not align were “wrong” and rejected.

The following is from *Formative Assessment and Science Education* (Bell & Cowie, 2001a), which provides numerous rich examples of assessment interactions in science classrooms, such as the following from a lesson on density. The researchers wrote it from field notes and transcripts:

The teacher started the lesson by reminding the students they had started to talk about density during the previous lesson. She asked for someone to tell her what it was. A student said it was ‘mass or volume.’ The teacher rephrased this as: ‘It is the mass of a certain volume.’ She emphasized that the certain volume was important and recorded on the board:

density = mass of a certain volume

She asked the student how they thought density, floating and sinking were linked. No one answered. She reminded them of the experiment in which they checked to see if cubes of different materials floated or sank. A student sitting beside her spoke to her.

The teacher said:

T. Z has a thought. Z?

Z: All those which weighted less than water floated and all those which weighted more than water sank.

The teacher restated Z's idea and asked if someone could put it in a sentence using density. A student offered an answer:

S: Whether or not something floats depends on its density

The teacher asked if someone could provide another phrase with 'tells us a little bit more?' A student discussed lead floating or sinking. The teacher asked for a general statement. A student offered more:

Things which are more dense that water sink and things that are less dense float.

The teacher asked the class:

Does that make sense to all of you? If not put your hands up.

There were nods and 'yeahs' around the room and she wrote on the board:

To compare the weight of materials we must use a fair test.

J gave the formula $\rho = m/v$

She continued writing on the board:

The rule of floating and sinking is

If it is denser than water_____

If is less dense it_____

She moved around the class and then added to the board:

If it is the same density it will_____

(Bell & Cowie, 2001a, pp. 104–106).

The researchers offer this as an example of planned and interactive formative assessment:

"...Her flexible plan allowed her to respond to student comments by appropriating their ideas and building on them, weaving them towards the learning she wanted. Hence, she undertook interactive formative assessment... This episode is viewed as an illustration of formative assessment in that the teacher provided the students with feedback on the appropriateness of their explanations." (p 106)

The first student's answer, "mass or volume," conveyed an uncertainty of meaning, probably not only of density but also of mass and volume. It was close to the correct *phrase*, however, that students had likely heard previously, and so the teacher could appropriate that contribution and turn it into "the mass of a certain volume." That appropriation, however, was of words only, since the meaning of these two phrases is quite distinct.

The teacher then posed the question to students of how density was related to floating and sinking, and "no one answered." After she reminded them of what they had seen, Z answered in terms of *weight*, and the teacher prompted for a restatement using the word density. Again, the emphasis was on the terminology and not on the substance of the ideas.

As in the previous examples, the researchers' analysis pays little attention to the substance of the students' reasoning, or to how the teacher attended to that substance. By this

age, the students had probably seen and heard the definition of *density* before, but there were several forms of evidence that they did not understand how density as an idea is distinct from weight or mass or volume. The teacher did not attend to that evidence; instead she attended—and showed them she was attending—to the terms they were using. The target content, then, appeared to have been the body of correct information, in the form of terminology.

Conceptualizing the target content in this way allows for clear definition of objectives. For this reason, it may be that the value placed on making learning goals explicit in lesson plans and as part of instruction (Heritage, 2010; Sadler, 1989) engenders a simplistic view of content as a body of information. By the same token, this conceptualization affords and is supported by standardization; thus standards are predominantly comprised of target information or propositional knowledge (Settlage & Meadows, 2002; Smith, Wiser, Anderson, & Krajcik, 2006; Van Sledright, 2010).⁵

Assessment Is Discussed in Terms of Particular Instructional Strategies, Techniques, and Procedures, Distinguishable From the Substance of Learning and Teaching

In a review of the formative assessment literature, Bennett (2011) describes the nature of the current debate around the definition of formative assessment. He argues that some, primarily from the testing industry, refer to it as an interim assessment and emphasize the actual instrument. Others, primarily educational researchers, place emphasis on the process of using data to inform instruction. Bennett suggests that the two are not so distinct, and argues that formative assessment involves a combination of task and instrument and process (Bennett, 2011).

Regardless, there is widespread and consistent acknowledgment that assessment should be an ongoing, inherent aspect of teaching and learning (Popham, 2008; Shepard, 2000; Gipps, 1994). Perhaps because of the emphasis on process, much of the literature focuses on particular strategies and techniques for teachers to employ to elicit student ideas. This is evident in the examples above, specifically in authors' reflections, which consider the teachers' actions—how they posed questions, listened to answers, provided feedback, and so on—but do little to consider the substance of those questions, answers, or feedback.

Here is one final example, from Shavelson's research group's (2008) multi-year project to include "curricular embedded formative assessment" in middle school science classrooms. The researchers had found no treatment effect on student achievement gains between control classrooms, where teachers taught the same lesson without the embedded assessment, and treatment classrooms, where teachers were enacting the embedded formative assessments (Shavelson, 2008). Working to understand these results, the group conducted a fidelity of implementation study that aimed "to determine whether the teachers implemented the critical aspects of the embedded assessments as prescribed by the *Guide* and to link the quality of implementation to the effectiveness of the formative assessments to improve student learning" (Furtak et al., 2008, pp. 374–375). Their analysis considered fidelity both to structural adherence of the assessment (implementation of all prompts, sequencing, placement of discussions, and timing) and to the quality of delivery, which focused on what they identified as critical "teaching strategies" designed to be consistent with formative assessment and science inquiry.

The following data table (see Figure 1) is from their analysis of a lesson on density. Students filled straws with different amounts of sand, sealed the straws, and tried floating them in water. We include here the final minute of a 4-minute transcript the researchers presented to illustrate their coding, the only segment in the episode that includes substantive

<i>Speaker/Dialogue</i>	<i>Coding Category</i>	<i>Codes Applied</i>
<i>Minute 4</i>		
Teacher: I'll show them to you quickly to refresh your memory. We had different amounts of sand in all the same sized straws. What did you think?	Focus of Instruction Student Task Making Students' Thinking Explicit	Whole Class Discussions Collecting student responses
Student: Because it weighed more.	Tracking Student Conceptions	Displaying students' ideas
Teacher: Okay, weighed more. Hm, I heard a lot of gasps. Because you probably felt me cringing. It's okay, Steven. And then I heard Steven go oh. Do you want to change that word?	Promoting Reasoning Argumentation	Asking students to elaborate Students respond to questions/statements posed by the teacher
Student: More mass.	Role of Evidence	Students cite Data/evidence from class
Teacher: Oh, thank you. I'm going to cross that out so Mrs. Jones can live another day and not have a heart attack. No, I think I know my own last name, guys. He's saying that straw four massed more. How do you know that?		
Student: Because it had the most sand in it.		

Figure 1. Coding table (Furtak et al., 2008).

student utterances. The teacher had been polling students over which straw they thought would “sink the furthest” in the water, and immediately prior to this minute had asked, “For those of you who thought that straw number four would sink the furthest, what were some reasons for that?” (Furtak et al., 2008, p. 374).

The student’s response that the straw “weighed more” apparently drew “gasps” from other students; the teacher spoke of “cringing,” and asked the student if s/he wanted to “change that word.” A student (it is not clear whether it was the same student) corrected the term to “mass.” The teacher jokingly emphasized the importance of the term and restated the idea, modeling a verb form: the straw “massed more.” Finally she asked, “how do you know that,” to which a student responded that it had the most sand in it.

As in the previous examples, there was little attention to disciplinary substance; again, the focus was on terminology. It was the *word* that made the teacher cringe, and the correction from “weight” to “mass” was sufficient in itself, without any explication of the difference in meaning.⁶

Here we highlight the complementary point, also evident in the previous examples: The analysis treats strategies without regard to substance. That is, the researchers were able to explain and justify the codings in their scheme without reference to students’ ideas. While this segment included codes of “displaying students ideas,” and “asking students to elaborate,” the analysis did not consider what, specifically, those ideas might have been. By “weighed more,” did the student mean “had more mass” or “was pulled harder downward?”

The researchers then analyzed the extent to which each teacher's implementation aligned with the intended delivery by quantifying the proportion of "critical strategies" to the total codes, and ranked teachers accordingly. They reported some correlation, although not a statistically significant one, between teacher ranking and students' achievement gains from pre-test to post-test. That is, the results failed to reject the null hypothesis that fidelity, construed as structural adherence, had no effect on achievement. This led the researchers to question their emphasis in professional development: "In hindsight, we believe that we should have put less effort into presenting teachers with many possible teaching strategies, and more effort into identifying what we believed were the most important strategies to help students learn. . ." (Furtak et al., 2008, p. 387).

The emphasis on strategies and techniques is widespread, from research to books for teachers (e.g., Black et al., 2003; Heritage, 2010; Keeley, 2008). The strategies are numerous and well intended, as they are designed to elicit rich and robust student thinking and reflection. However, as is evident in all four of the examples we have presented, the focus on instructional strategies can undermine everybody's attention to the *very ideas those strategies were supposed to make visible*. In other words, if researchers and teachers are watching to see when and whether teachers ask students to elaborate, they may not attend to what the students actually have to say.

Relevant Research on Learning

The literature in assessment, we have argued, presumes traditional notions of content as a body of correct information. Research on learning, of course, has long challenged such notions. Readers of this journal are familiar with the central arguments, so we limit our discussion to brief comments regarding conceptual change and intuitive epistemologies.

Research on Conceptual Change

While there is little attention to the substance of student thinking in research on assessment, some articles do cite research on student conceptions as background for their work.

Morrison and Lederman (2003), for example, were centrally concerned with teachers' diagnosing "preconceptions": "research has established that students enter their science classrooms with ideas about the natural world that are not in alignment with accepted scientific beliefs" (p. 849). They were concerned that teachers become aware of the "depth and tenacity of the students' preexisting knowledge." They found that only one teacher of the four they studied, Bill, worked to diagnose preconceptions and showed "a thorough knowledge of the many preconceptions students bring to a physics class":

Bill talked about how these reoccurring preconceptions come from students' experiences and also from elementary school teaching. He talked about gravity and how many students he had taught both in high school and college "get it stuck in their mind that the bigger object has to fall faster than a smaller object and an object going horizontal falls at a different rate than something just straight down." He also discussed how electricity and circuitry always were an area where students had preconceptions. Bill was asked if he used this knowledge of the reoccurring, common preconceptions when he taught these concepts. He replied that he did try to describe the common preconceptions and explain the correct concepts (Morrison & Lederman, 2003, p. 859).

This reflects a widely subscribed view of misconceptions, or preconceptions, as incorrect conceptions students hold that are contrary to the instructional objectives.

Much of the research on misconceptions, however, was written to challenge the idea that it is sufficient or even of primary importance to “explain the correct concepts.” In particular, researchers meant to highlight the *rationality* of students’ prior conceptions. Strike and Posner (1992), motivated largely by what they saw as misuses of their earlier work (Posner, Strike, Hewson, & Gertzog, 1982), argued pointedly, “If conceptual change theory suggests anything about instruction, it is that the handles to effective instruction are to be found in persistent attention to the argument and in less attention to right answers” (p. 171).

Moreover, there has been progress in models beyond views of students’ holding or not holding conceptions. Strike and Posner (1992) argued, “it is very likely wrong to assume that misconceptions are always there in developed or articulated form. . . . Misconceptions may be weakly formed, need not be symbolically represented, and may not even be formed prior to instruction” (p. 158). Others, notably diSessa (1993), have challenged accounts that attribute coherence, “depth and tenacity” to prior conceptions, giving evidence of contextual sensitivity and fragmentation to student thinking. These authors have raised empirical and theoretical reasons to doubt the view of prior conceptions as obstacles to learning (Nobes et al., 2003; Smith, diSessa, & Rochelle, 1993; Taber, 2000), although this continues to be the subject of some debate.

Research on Epistemologies

Further research on learning focuses on students’ epistemologies, that is, on how students understand what knowledge, reasoning and learning entail, in science or science class. Data from interviews, observations, and survey instruments have provided evidence of students’ thinking of knowledge in science as piecemeal information that may not connect to everyday experience, to be accepted on the authority of the teacher or text (Hammer, 1994; Smith, Maclin, Houghton, & Hennessey, 2000). Research on students’ understanding the “Nature of Science” has been mainly concerned with knowledge about science as a formal discipline, connecting to issues in science, technology and society, while research on epistemologies is more concerned with how students’ experience and understand their own knowledge and learning. The literatures overlap, and there are interesting questions regarding how the two relate students’ personal epistemologies and their sense of professional science (Driver, Leach, Millar, & Scott, 1996; Hogan, 2000; Sandoval, 2005).

Here we note only that matters of assessment figure centrally throughout this work, in two respects. First, this research directly concerns students’ sense of the assessment of ideas. How do they, as learners, assess whether an idea they are considering has merit, whether they should accept it as true, whether they understand it? How do they expect that assessment happens within science? Second, it speaks to how experiences in science class indirectly “tell” students about assessment. By and large, ideas in science class are, as a matter of pedagogical practice, assessed for quality by reference to the canon of “currently accepted scientific knowledge”—in tension if not open conflict with practices of assessment in the discipline, where the esthetic, at least, holds that “the authority of a thousand is not worth the humble reasoning of a single individual” (Galileo, 1632).

To be sure, assessment is part of the substance we hope students to learn (Coffey, 2003a; Coffey & Hammer, in preparation; Duschl, 2008), and this adds complexity to the challenges of formative assessment. This was part of Strike’s and Posner’s argument for “persistent attention to the argument” in science class, that students should come to understand science, including the assessment of ideas within science, as rational. A problem they hoped to address is that students’ everyday experience in science class was not of rational argument. They wrote, “Thus, occasional epistemological sermons (or courses in philosophy of science)

should not be expected to enhance a student's conception of science as a rational activity if the student's prior experience and current outlook are at odds with them" (Strike & Posner, 1992, p. 171).

Still further work concerns how students *frame* (Goffman, 1974) what is taking place in science class. A variety of accounts have drawn distinctions between students, on the one hand, participating as nascent scientists, trying to make sense of the world as scientists do, and, on the other hand, performing obligations in exchange for credit in the classroom. Lemke called the latter "playing the classroom game"; others have referred to it as "doing the lesson" (Jimenez-Alexiandre, Rodriguez, & Duschl, 2000), "doing school" (Pope, 2001), and being "obedient" (Tobias, 1993). Research on framing, and more specifically *epistemological framing* (Hammer, Elby, Scherr, & Redish, 2005; Redish, 2004), discusses the dynamics of how participants form a sense of what is taking place, including with respect to "meta-communicative messages" they exchange over what they think is happening. The teachers' moves in each of the examples above, focused on terminology, are examples of such messages. They communicate to students how the teachers frame what is taking place.

In these ways, research on conceptual change and research on epistemologies argue against traditional notions of content, along with assessment practices that center on comparing student statements to the predefined target information. None of this is to argue against formative assessment; it is to argue for developing practices of formative assessment more aligned with and reflecting practices of assessment in science. Ideas and reasoning should be "good" in science class in similar ways that they are "good" in science: They make sense; they are supported by the available evidence; they have explanatory and predictive power.

Assessment as Attending to Disciplinary Substance

Discussions about formative assessment are fundamentally about teachers' awareness, in order that they can make adjustments that respond to students' reasoning and participation. To this point, we have raised concerns over how the literature has focused on strategies for assessment but taken disciplinary substance for granted. We have shown that prominent examples in the literature have tacitly treated subject matter as information, and the assessment of student thinking as a check on its correctness against that information.

Effective assessment in science education, we argue, should involve genuine, extended attention to the substance of student reasoning, on at least two levels. Teachers should elicit and pay "persistent attention" (Strike and Posner, 1992) to students' arguments. What reasons do students have for answering as they do? What evidence and logic are they using? In this, the teachers are not only becoming aware of student reasoning but modeling for students how they should focus their attention in science. In other words, they are assessing student reasoning in ways that are consistent with how students should learn to assess ideas as participants in science.

At another level, formative assessment should involve awareness of how students are engaging in disciplinary practices. Are students reasoning about the natural world, or are they focused on what they are "supposed to say," playing the "classroom game" (Lemke, 1990) of telling the teacher what they think she wants to hear? We see these multiple and interacting goals highlighted in other areas of literature.

Practices of Formative Assessment Elsewhere in the Literature

We are arguing that formative assessment should be understood as a matter of attention to disciplinary substance, and in this sense it should be inherent throughout classroom

activity, not restricted to specifically designate “assessment activities.” While this is often in the rhetoric of formative assessment, we have argued, it is often not achieved. There are, however, models of formative assessment in work that identifies itself in other ways.

A prime example is Ball’s (1993) account of “dilemmas” teachers face in honoring children’s reasoning while at the same time guiding them toward disciplinary understanding and practices. When the student Sean in her third grade class claimed that the number 6 could be both even and odd, Ball’s response was to elicit his argument, pay it attention, and guide other students to pay it attention as well. This led to the discovery and definition of a class of numbers they called “Sean numbers” that had the interesting property that they “have an odd number of groups of two,” which allowed further investigation into patterns and properties.

Ball never described what she or her students were doing in class as assessment, but that is precisely what it was: They were assessing Sean’s idea, its validity, merits, weaknesses, and possibilities; they were working with the idea as nascent mathematicians. Ball was assessing at another level, as well, in her attention to the students’ participation in mathematical inquiry, to the community they were forming and the understandings they were developing of what it means to engage with mathematical ideas.

There are many other accounts in the literature that, like Ball’s, depict close attention and responsiveness to student thinking (Franke and Kazemi, 2004; Lampert, 2001; Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001; Warren and Rosebery, 1995). Hammer (1997) described the “tensions” among multiple objectives and multiple dimensions of awareness of student reasoning and participation in his high school physics class. In one episode, for example, a group of students devised their own experiment and collected their own evidence to arrive at the conclusion that StyrofoamTM conducts electricity. It was a group that included students who had generally been reluctant to pursue their own ideas, and Hammer wanted to support their initiative and nascent scientific stance, but this was in tension with their learning which materials do and do not conduct electricity. (Styrofoam does not.) Hammer described his own and others’ accounts of their practices as “discovery teaching,” in which “successful instruction depends on teachers’ often unanticipated perceptions and insights” into student thinking and participation.

In a study of middle school math teachers responsiveness to student ideas, Pierson (2008) defined “responsiveness” as “the extent to which teachers ‘take up’ students’ thinking and focus on student ideas in their moment-to-moment interactions,” and in particular “High II” responsiveness in which the focus is on the students’ meaning and logic for the immediate purpose of understanding their reasoning on its own terms. She distinguished High II from High I responsiveness, in which the teacher worked to identify student ideas with the purpose of correcting them. In Pierson’s language, our critique of the formative assessment literature is that there is evidence of “High I” but not of “High II” responsiveness.

With data from 13 teachers, Pierson found a strong, significant correlation between High II responsiveness and student learning, but not with High I. Working to explain this finding, she discussed how “discussions high in responsiveness can act as formative assessments, which research indicates is positively related to student learning,” citing work from the formative assessment literature. She also cited work by Saxe, Gearhart, and Seltzer (1999) who found that “integrated assessment,” which they defined as “the extent of opportunity for students to reveal their understandings, to receive interpretations of their contributions, and to provide interpretations of others’ contributions,” (p. 11) correlated with students’ achievement in problem solving.

Not all of these accounts speak of formative assessment, but that is what they depict: teachers continuously attending to the students' understanding, reasoning, and participation. They do not focus on particular, discrete strategies; rather, they show teachers adopting stances of "respecting students as thinkers," as Ball put it. They learn about what and how students are thinking and participating, and they use that information to guide their instruction.

While we expect readers are familiar with some of this work, at least Ball's (1993), it will be helpful to have an example of what we mean by this continuous attention to students' reasoning and participation. We have chosen one from the classroom of a teacher who, like the teachers in the examples above, was working to reform his practices.

An Example of Substantive Formative Assessment

The following example comes from a corpus of data collected during a 3-year study⁷ that examined what high school science teachers attended to in their classroom interactions with students and how that informed their instructional decision making and curricular modifications. Terry Grant, a co-author of this article, was one of 28 collaborating teachers. The participating high school science teachers worked in subject-matter cohorts of 8–10 others and university-based researchers. For 3 years, they met bi-weekly for 2 hours to share and discuss video and student work from their classroom teaching. In addition, all of the project teachers met together each summer for a 1–2 weeks workshop. The focus of the regular meetings was discussion of student ideas and reasoning, along with consideration of possible next moves. Teachers, often in collaboration with university-based researchers, wrote case studies about their student learning and their instructional decision-making, grounded in their video data, student work, and their personal notes.

The excerpt below comes from Terry's case study of a class he taught in the 2nd year in the project, and his 10th year as a classroom teacher.⁸ It was also his first year at a new school, and he decided to use that opportunity to remake his practices, specifically to become more aware and responsive to his students' thinking.

Terry's 9th grade biology class was scheduled to begin a unit on the chemistry of life, which assumed some basic understandings of matter, atoms and molecules. He had intended a 15-minute review of these concepts, starting with the textbook definition they had read: "Matter is anything that occupies space and has mass." He began by asking students to say what it means to take up space and have mass. They did this with some hesitancy; as Terry pressed they gingerly offered that you "can see it" and "can touch it," that it "weighs something."

Terry thought to remind students about the difference between mass and weight, which they had studied.⁹

Teacher: What influences your weight? Do you weigh more on the earth or on the moon?

Barb: Gravity!

Teacher: Ohhh (quietly). And. So what's the difference between your weight and your mass. Standing right here. [silence for several seconds] Nothing.

Terry decided to set the topic aside, saying, "I don't think it's going to be that significant" for the topic at hand, and for the rest of the conversation, he let them say "weight." Terry asked about a table, which the students all thought is matter, and then water. Barb answered, "No."

Teacher: Why?

Barb: I think it's composed of molecules.

Teacher: OK, which are?

Barb: Matter? [barely audible]

Teacher: Are they? Perhaps it doesn't matter, but you kind of went, "Matter?" [mimicing the tentative tone]

[quiet laughter]

Barb: Yes. [clearly stated]

Terry guided them again to apply the definition of matter to water, and students quickly agreed: It has weight, and it takes up space, such as in oceans. One student offered, "doesn't matter have to do with a state? Like liquid and solid and gasses?" Since water is one of those, it must be matter. Terry then asked whether air is matter, and many voices in unison said "No." Barb and Brianna said "You can't weight it." and "You can't see it." Several students were speaking at once, and then Barb had the floor.

Barb: It takes up space but you can't feel it, Like you can't bump into it. . . Cause air is everywhere, except for in water. . . well actually no there is air in water.

Terry: Adria. Shh. Adria, did you have your hand up?

Adria: I was gonna say you could feel it. Or you can't feel it.

Barb: Yeah, well you can feel wind. [Overlapping talk.]

Terry: Yeah, What's wind?

Brianna: Air. . . air blowing [Overlapping talk.]

Terry: So can we weigh it?

Students: No. . . no

Terry: Those are the issues we've got to resolve. Can we weigh it?

Maggie: No [Multiple students. Emphatic.]

Terry: How could I weigh it? What could you do to weigh it?

Barb: You could like put it in a balloon or something but there's the weight of the balloon so you couldn't weigh it.

Terry: I haven't got a scale with me today. (Walks over to his desk and pulls out a bag of balloons). . . So I have balloons, right? (Tosses Barb a balloon.) Blow it up.

By this point, the class was past the 15 minutes Terry had planned. We pause here to consider what is happening in this snippet.

First, it is immediately obvious that the students showed hesitancy and confusion over the concept of *matter*: They had trouble articulating what it meant; they were not sure at first whether water is matter and they remain unsure about air.

Less obvious, but as important, is how the students approached the topic. Barb, the most outspoken student in the room, seemed to be trying to find the right terminology ("gravity,"

“molecules”), and her questioning “Matter?” suggested that she was “playing the classroom game” (Lemke, 1990). Another student reasoned in terms of material they had covered. Her “doesn’t matter have to do with a state” sounds like a comment based on the structure of a curriculum rather than on the substance of ideas.

Terry noticed and responded at both levels. For one, much like the examples we presented earlier from the assessment literature, he discovered that material he thought would need only a quick review was deeply problematic for the students, first the definition of matter and, within that, the difference between mass and weight. He posed questions, much as in other examples, to find out what they did and did not understand.

At the same time, he noticed and considered how they framed what was taking place. Terry chose to set the question of mass versus weight aside, rather than give an explanation on the distinction (which they had “covered”), and he allowed them to continue to use the (formally incorrect) term *weight* rather than *mass*, because he wanted to encourage their genuine engagement with the ideas. He remarked on Barb’s uncertainty, when she asked “matter?” And he focused on eliciting their reasoning. He confirmed that solid objects are matter, and that water is matter, but not until the students seemed to reach their own consensus on these points.

With the class divided on whether air is matter, he made them responsible for the “issues we need to resolve,” taking his lead from them, with one exception: When he asked “can we weigh it?,” the students said no, but he proceeded anyway, asking “how can we weigh it?” If that was a lapse, possibly he was trying to press for a resolution, to allow them to move on with the plan for the day. Still, his question did not convey his view on whether air is matter.

Terry had the class continue this new inquiry on a new, unplanned topic. Like Ball hearing Shea’s idea that six is an odd number, Terry changed his goals for the lesson in response to what he heard in their thinking. We present one more snippet, to highlight more of the formative assessment inherent in Terry’s close attention to the students’ reasoning.

As Barb inflated the balloon, Mikela commented “that one’s stretching because she’s blowing air into it,” and Terry asked if that meant there was a different “amount of stuff in it.” Students spoke over each other, some saying “yes” and some “no”; among the remarks was Lauren’s: “Air and matter is closed up.” Barb said that air was “occupying space in there,” referring to the balloon, and Terry asked if air is “occupying space in the room.” Again, students spoke over each other, giving a mix of answers.

Terry: Are you saying? And I’m asking, I’m not telling. That it takes up space when it’s in HERE [in the balloon], but it doesn’t take up space when it’s in the room.

India: No.

Terry: Is that the general consensus?

Barb: No! Actually that’s right cause you can’t put something inside that balloon with air in it.

[Several students speaking at once]

Terry: OK. What would happen to the air in the balloon, if I put water in it too?

Barb: There wouldn’t be as much air.

Terry: Because?

India: The water’s taking up space.

Terry: Okay. What would happen to the air in the balloon, if I put water in it too?

Barb: There wouldn't be as MUCH air.

Terry: Because?

India: The water's taking up space.

Terry: OK, so...

Laura: The air is the space.

Terry: Say it again.

Laura: The air IS the space.

Terry: So air IS the space. Are you saying it takes up space? Is that the idea?

Ari: The air is the space that gets taken up.

Terry: So it's an empty space until I put water in it? I'm trying, I'm trying to work your way... I'm not trying to say you're right or wrong, I'm asking. This is not a graded assignment or anything.

Ari: Yes.

Terry: Yes? How many people agree with that? Air is empty space that the water is going to take up when I pour water in. If I were more daring I would've brought a couple of water balloons too. I'm afraid they'll blow up in here... So think about this, some of you have this look on your face like "I don't know for sure," is this just empty space which we filled up with water, or is there something in there?

Brianna and Laura (simultaneously): There's something in there.

Terry: Okay, what's the something?

Students: Air! Air!

Terry: So, does it take up space?

Students: Yes!

Laura: I'm confused!

India: Oh my god!

Terry: You don't sound convinced, you're giving me "ummmm." Yea, go ahead.

India: But when, when something else goes in there, doesn't some of the air leave?

Later in the period the class tried to compare the weights of an inflated and empty balloon, found no difference, and discussed what that meant, including the possibility that the scale was not sufficiently sensitive. The next day, they tried again, with a more sensitive scale, which showed the inflated balloon a bit heavier. When a student raised the concern that blowing up a balloon would contaminate the air with saliva, another student suggested using an air pump from the gym, which Terry let her fetch. In the end, the class concluded that air is matter.

Of course there are a number of issues we could consider about this class: Is it practical to take so much time on a "review" topic? Was the conversation too dominated by a few

individuals? What might Terry have done differently to address the students' needs? These are important issues, but our focus here is on Terry's attention to the disciplinary substance of his students' thinking as an example of formative assessment.

Again, we begin by considering the students' thinking. For many, seeing the balloon inflate showed air taking up space. Barb and India argued that, if there were water in the balloon, then there would have to be less air, because the water would be taking up some of the space. For others, though, the idea that air takes up space seemed to be in tension with their intuition that air does not take up space in the room. Lauren and Ari gave a clear articulation to a very different way of thinking, that air and space are the same thing.

Along with the evidence of their conceptual understanding, there was evidence of how they were approaching the topic. Rather than looking to Terry or the textbook, in this snippet, and rather than quoting terminology, the students were arguing on their own terms based on their own reasoning and observations. The conversation was lively and robust, with more students entering the fray.

Terry was attending—and responding—at both levels, trying both to understand their reasoning and to assure them that was what he wanted to be doing. He emphasized that he was “asking... not telling,” in an energetic interrogation of their thinking. When Lauren spoke up with her very different view, it was not clear to Terry what she meant, and he worked to understand, telling her “I’m trying to work your way” and, “I’m not trying to say you’re right or wrong.” Ari’s clarification seemed to help, and Terry threw the idea back to the class, to find out how many others thought similarly, remarking on students’ uncertain faces. Laura had changed her mind, apparently, but still seemed torn, and despite the overwhelming “yes” from students that air is “something” and that it takes up space, he tried to keep the question open, making room for students to continue to question.

From Strategies to Attention. As an example of formative assessment, this account of Terry’s class is similar in several respects to the examples we quoted from the literature: He posed questions to students, listened to their answers, and what they had to say informed how he moved forward. That is, if we use the literature for guidance, we could support the claim that this is formative assessment simply by considering what Terry was doing.

However, we argue, it is not sufficient to consider only the teacher’s actions. The core of formative assessment lies not in what teachers *do* but in what they *see*. The point is teachers’ *awareness and understanding* of the students’ understandings and progress; that’s what the strategies are for. To appreciate the quality of a teacher’s awareness, it is essential to consider disciplinary substance: What is happening in the class, and of that, what does the teacher notice and consider?

In our critique of the examples from the literature, we argued that they (1) neglected the disciplinary substance of student thinking, (2) presumed traditional targets of science as a body of information, selected in advance, (3) treated assessment as strategies and techniques for teachers. In our presentation and analysis of Terry’s class, we worked to do something different in each of these respects.

First, as we have discussed, we began with student thinking in our analysis. Second, like Terry, we considered student thinking not only with respect to its alignment with the canonical ideas but also with respect to the nature of the students’ participation. Students’ acceptance that air is matter (or that plants feed by photosynthesis or that sunspots are magnetic phenomena, etc.) could be seen as alignment with the canonical ideas. However, if students accept those ideas on the teacher’s authority, rather than because they see them supported over other ideas by evidence and reasoning, then they are at odds with the practices of

science. For this reason, it is essential that formative assessment—and accounts of it in the literature—consider more in student thinking than the “gap” (Black & Wiliam, 1998a; Sadler, 1989) between student thinking and the correct concepts.

Moreover, it was Terry’s attention to the disciplinary substance of student thinking that led him to abandon his original plan for the lesson. Formative assessment *created* objectives for him that he did not have at the outset, and again at two levels. One objective was conceptual, that students understand the concept of *matter*. Another was at the level of how students approach the topic, and there we could see Terry working to move students into engaging the material as nascent scientists, and away from the “classroom game” (Lemke, 1990) of telling him what they think he wants to hear.

To the third point, it is not possible to distinguish any particular strategies from the activity as a whole. Conceptualizing assessment as attention, Terry was formatively assessing student thinking by closely attending to it. He wanted to understand what they were thinking and why, as would any participant in any meaningful discussion. Formative assessment should be understood and presented as nothing other than genuine engagement with ideas, which includes being responsive to them and using them to inform next moves.

Terry’s formative assessment was continuous with what he hoped students to learn: practices of assessing the quality of ideas for their fit with experience and reasoning. Effective assessment is part of the substance students should learn. It is important that they understand whether air is matter; it is also important that they understand what goes into deciding whether air is matter, and that, fundamentally, is the assessment of an idea. Thus the students were learning to assess ideas as nascent scientists, rather than as compliant students. Understanding these discipline-based assessment criteria is part of what educators should help students learn. As students learn to engage in disciplinary assessment, they are learning a fundamental aspect of science (Coffey, 2003a; Coffey & Hammer, in preparation; Duschl, 2008).

Reframing Assessment

We have argued that, in focusing attention to strategies and techniques, the literature on formative assessment has generally presumed traditional notions of disciplinary content as a body of information. These notions have shaped the filters and criteria for what counts as disciplinary substance in student thinking, and so we see teachers and researchers attending to how students’ thinking aligns with the target information, with an emphasis on terminology, more than to the meanings students are trying to convey or to the rationality of their reasoning. In this, we have argued, the literature is at odds with research on learning, and it is at odds with disciplinary practices. If assessment criteria are incongruous with what happens in the discipline, educators can misconstrue what counts within disciplinary activities and distort for students what engaging in science activities entails.

There are, clearly, a variety of reasons for this state of affairs, including the ways *summative* assessments are constructed and valued in the much larger educational system. It is only natural that practices of formative assessment will tune to support desired outcomes. Much, certainly, has been discussed in the literature over the influence of high-stakes standardized testing (e.g., Valli, Croninger, Chambliss, Graeber, & Buese, 2008).

In closing this article, we consider another possible reason for the literature’s emphasis: Researchers and teacher educators seem to believe that *strategies* are what teachers need first or most, to help them engage in formative assessment. At least two of the research projects cited above (Black et al., 2003; Shavelson, 2008) explicitly organized their professional development efforts around the premise that enacting well-defined assessment strategies will

elicit and facilitate teachers' awareness of student understanding, awareness that is difficult to achieve.

Discussing on-the-fly assessment that "occurs when 'teachable moments' unexpectedly arise in the classroom" (p. 4), Shavelson (2006) wrote:

Such formative assessment and pedagogical action ("feedback") is difficult to teach. Identification of these moments is initially intuitive and then later based on cumulative wisdom of practice. In addition, even if a teacher is able to identify the moment, she may not have the necessary pedagogical techniques or content knowledge to sufficiently challenge and respond to the students. (p. 4)

There is support for this position in research on teacher "noticing" (Sherin & Han, 2004), in findings that most pre-service and inservice teachers have difficulties noticing student thinking (Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Sherin & Han, 2004). Other research on formative assessment has argued that, although teachers can often make reasonable inferences about student understanding, they face difficulties in making "appropriate" instructional moves (Heritage, Kim, Vendlinski, & Herman, 2007).

From these perspectives, the examples we gave earlier of responsive teaching reflect the work of experienced, accomplished practitioners. There is little hope of recruiting or training a million teachers like Deborah Ball, the reasoning goes, but it is not difficult to imagine large-scale implementation of well-defined strategies, such as "traffic-lighting" (Black et al., 2003) "two stars and a wish" (Keeley, 2008), reflective toss (van Zee & Minstrell, 1997), and wait time (Rowe, 1974). These represent clear, tangible steps teachers can take in class, and so research on formative assessment has produced professional development to leverage such strategies (Black, Harrison, Lee, Marshall, & Wiliam, 2004a; Furtak et al., 2008).

The work we have done in the first half of this paper belies that premise. Examining particular instances in four articles from prominent work, we showed that, while the teachers were using the strategies they had been taught, they were not engaging with student ideas. The example we provided in the second half of this paper, in contrast, shows a teacher's early success in becoming more aware of and responsive to student thinking, without the benefit of any particular strategies. Of course, Terry was an experienced teacher, but his entrance into these practices did not begin with strategies. It began, rather, with a shift of *attention*, with a shift in how he framed, and asked his students to frame, what was taking place in class.

An orientation towards responsiveness to students' ideas and practices resonates with work in teacher education, particularly in mathematics, that has pushed for more practice-based accounts of effective preparation (Ball, Thames, & Phelps, 2008; Kazemi, Franke, & Lampert, 2010), and that comes with calls for learning to teach "in response to what students do" (Kazemi, Franke, & Lampert, 2009; p. 1) and more attention to "demands of opening up to learners ideas and practices connected to specific subject matter" (Ball & Forzani, 2011; p. 46).

We challenge the view that it is difficult for teachers to learn to attend to the substance of student thinking. Recent work in science and math teacher education (Coffey, Edwards & Finkelstein, 2010; Kazemi et al., 2009; Levin, Hammer, & Coffey, 2009; Levin & Richards, 2010; Singer-Gabella et al., 2009; Windschitl, Thompson, & Braaten, 2011), has presented evidence of novice teachers' attention to student thinking, novices whose preparation emphasized awareness and interpretation of student thinking as evident in video records and written work. By this reasoning, much depends on how teachers frame what they are doing, and a primary emphasis on strategies may be part of the problem. Assignments that direct teachers

and teachers-in training to *what they are doing* may inhibit their attending to *what students are thinking*. Our analyses above show that the same applies to researchers.

Thus, in closing, we argue for a shift of researchers' attention to *attention*, that is from the strategies teachers use to the focus of their attention in class, and with that a re-framing of what assessment activities entails. First and foremost, we propose that it is essential for teachers to frame what is taking place in class as about students' ideas and reasoning, nascent in the discipline. Formative assessment, then, becomes about engaging with and responding to the substance of those ideas and reasoning, assessing with discipline-relevant criteria, and, from ideas, recognizing possibilities along the disciplinary horizon. Framed as such, assessment demands attending to substance, in research, professional development, as well as in classrooms. With this reframing, many teachers will be able to do something much more akin to what Ball was doing, provide systemic support for that focus of their attention to substance. As the example from Terry's classroom illustrates, with attention on substance and a framing of making sense of and responding to ideas, formative assessment moves out of strategies and into classroom interaction, with roots in disciplinary activity and goals.

Notes

¹All of the citation counts are from Google Scholar, as of July 10, 2011.

²Beyond this particular instance, Black and Wiliam's influence on the work of formative assessment has been profound, again as evident in the citations of their work—2,476 for Black and Wiliam (1998), with an additional 1,910 for the pamphlet form.

³Bell and Cowie's (2001a) article published in *Science Education* included one more abbreviated example, and refers out to their related book for more extended examples. For purposes of this argument, we selected the example from the book because of the detail from classroom transcript, researcher's notes, and teacher reflection.

⁴In a sense, this is a rough description of the modern understanding: The sun is "gas ball" that's "on fire," although a different kind of "fire" than from everyday experience.

⁵A community of researchers and educators interested in articulating learning progressions is engaged in a related conversation, as the community struggles to construct and conceptualize learning progressions, beyond canonical conceptual attainments (e.g., Alonzo & Gotwals, in press; Sikorski & Hammer, 2010; Shepard, 2009). While some have made a compelling case for the promise of learning progressions for purposes of formative assessment (Alonzo & Steedle, 2009; Shepard, 2009), that topic is beyond the scope of this paper.

⁶*Weight* is the more directly relevant idea: The straw with more sand sinks further because the downward force on it by the earth is greater. Another mechanism for increasing the downward force (e.g., magnets) would have the same effect. But neither the teacher nor the researchers attended to either the student's or the disciplinary meaning of the explanation.

⁷NSF ESI 0455711.

⁸This excerpt comes from a fuller case that is part of a collection of video and written cases of high school science teaching and learning (Levin, Hammer, Elby, & Coffey, in press).

⁹In this instance, we note, *mass* is the appropriate concept. An object is matter because it has mass, not because there is a force on it by the earth.

This work was supported in part by the National Science Foundation, under grants ESI 0455711 and DRL 0732233. The views expressed herein are those of the authors and not necessarily those of the NSF.

References

- Alonzo, A., & Gotwals, A. W. (Eds.). (in press). *Learning progressions in science*. Rotterdam, The Netherlands: Sense Publishers.
- Alonzo, A., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education* 93, 389–421.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, 93(4), 373–397.
- Ball, D. L., & Forzani, F. M. (2011). Teaching skillful teaching. *The Effective Educator*, 68(4), 40–45.
- Ball, D. L., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Bell, B., & Cowie, B. (2001a). *Formative assessment and science education*. Netherlands: Kluwer Academic Publishers.
- Bell, B., & Cowie, B. (2001b). The characteristics of formative assessment in science education. *Science Education*, 85, 536–553.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, and Practice*, 18(1), 5–25.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. London: Open University Press.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004a). *Working inside the black box: Assessment for learning in the classroom*. London: Kings College.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004b). *Working inside the black box: Assessment for learning in the classroom*. *Phi Delta Kappan*, 86(1), 8–21.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology*, 79(4), 474–482.
- Butler, R., & Neuman, O. (1995). Effects of task and ego-achievement goals on help-seeking behaviours and attitudes. *Journal of Educational Psychology*, 87(2), 261–271.
- Coffey, J. E. (2003a). *Making connections: Engaging students in assessment*. Unpublished Doctoral Dissertation. Stanford University, Stanford, CA.
- Coffey, J. E. (2003b). Involving students in assessment. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom*. Arlington, VA: National Science Teachers Association Press.
- Coffey, J. E., & Hammer, D. (in preparation). *Engaging Students in Disciplinary Assessment*.
- Coffey, J. E., Edwards, A. R., & Finkelstein, C. (2010). Dynamics of disciplinary understandings and practices of attending to student thinking in elementary teacher education. *Proceedings of the 9th International Conference of the Learning Sciences*.
- diSessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10(2–3), 105–225.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.
- Duschl, R. (2008). Science education in three part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 28(1), 268–291.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist Special issue. Psychological Science and Education*, 41(10), 1040–1048.

- Erickson, F. (2007). Some thoughts on “proximal” formative assessment of student learning. In G. D. Fenstermacher (Ed.), *Yearbook of the National Society for the Study of Education* (pp. 186–216). Malden, MA: Blackwell Publishing.
- Franke, M. L., & Kazemi, E. (2004). Learning to teach mathematics: Focus on student thinking. *Theory into Practice*, 40(2), 102–109.
- Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the fidelity of implementing embedded formative assessment and its relation to student learning. *Applied Measurement in Education*, 21(4), 360–389.
- Galileo, G. (1632). *Dialog on the two chief systems of the world*. Translated by Drake, Stillman (2001). New York: Modern Library.
- Gipps, C. (1994). *Beyond testing: Toward a theory of educational assessment*. London: Falmer Press.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York: Harper and Row.
- Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, 12, 151–183.
- Hammer, D. (1997). Discovery teaching, discovery learning. *Cognition and Instruction*, 15(4), 485–529.
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of Learning: Research and Perspectives* (pp. 89–120). Greenwich, CT: Information Age Publishing.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues Practice*, 28(3), 24–31.
- Hogan, K. (2000). Exploring a process view of student knowledge about the nature of science. *Science Education*, 84(1), 51–70.
- Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children’s algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, 38(3), 258–288.
- Jimenez-Alexiandre, M. P., Rodriguez, A. B., & Duschl, R. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, 84, 757–792.
- Jos, E. (1985). The right question at the right time. In H. Wynne (Ed.), *Primary science... Taking the plunge*. Oxford: Heinemann.
- Kazemi, E., Franke, M., & Lampert, M. (2009). Developing Pedagogies in Teacher Education to Support Novice Teachers’ Ability to Enact Ambitious Instruction. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing Divides, Proceedings of the 32nd Annual Conference of The Mathematics Education Research Group of Australasia*, Vol 1, pp. 11–29. Palmerston North, New Zealand: Mathematics Education Research Group of Australasia.
- Keeley, P. (2008). *Science formative assessment: 75 practical strategies for linking assessment, instruction, and student learning*. Arlington, VA: National Science Teacher Association.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven: Yale University Press.
- Lemke, J. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex Publishing Corporation.
- Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers’ attention to student thinking. *Journal of Teacher Education*, 60(2), 142–154.

Levin, D. M., Hammer, D., Elby, A., & Coffey, J. E. (in press). Educating the responsive practitioner: Case studies in secondary science teaching. Arlington, VA: National Science Teachers Association Press.

Levin, D. M., & Richards, J. (2010). Exploring how novice teachers learn to attend to students' thinking in analyzing case studies of classroom teaching and learning. *Proceedings of the 9th International Conference of the Learning Sciences*.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Minstrell, J., & van Zee, E. H. (Eds.). (2000). *Inquiring into inquiry learning and teaching in science*. Washington, DC: American Association for the Advancement of Science.

Minstrell, J., & van Zee, E. (2003). Using questioning to assess and foster student thinking. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom*. Arlington, VA: National Science Teachers Association Press.

Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understandings of students' preconceptions. *Science Education*, 87, 849–867.

National Research Council. (2001a). *Classroom assessment and the national science education standards*. Washington, DC: National Academies Press.

National Research Council. (2001b). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Nobes, G., Moore, D. G., Martin, A. E., Clifford, B. R., Butterworth, G., Panagiotaki, G., Siegal, M. (2003). Children's understanding of the earth in a multicultural community: Mental models or fragments of knowledge? *Developmental Science*, 6(1), 72–85.

Nystrand, M., Wu, L. L., Gamoran, A., Zeisler, S., & Long, D. A. (2003). Questions in time: Investigating the structure and discourse. *Discourse Processes*, 35(2), 135–198.

Pierson, J. (2008). *The relationship between patterns of classroom discourse and mathematics learning*. Unpublished Doctoral Dissertation, University of Texas at Austin.

Pope, D. (2001). "Doing School": How we are creating a generation of stressed out, materialistic and miseducated students. Yale: Yale University Press.

Popham, W. J. (2006). Phony formative assessments: Buyer beware. *Educational Leadership*, 64(3), 86–87.

Popham, J. W., (2008) *Transformative assessment*. Alexandria, VA: Association of Supervision and Curriculum Development.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science*, 28(1), 4–13.

Redish, E. F. (2004). A theoretical framework for physics education research: Modeling student thinking. In E. F. Redish, C. Tarsitani, & M. Vincentini (Eds.), *Proceedings of the Enrico Fermi Summer School, Course CLVI* (pp. 1–63). Bologna: Italian Physical Society.

Rowe, M. B. (1974). Reflections on wait-time: Some methodological questions. *Journal of Research in Science Teaching*, 11(3), 263–279.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89(4), 345–372.

Saxe, G., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction*, 17(1), 1–24.

Settlage, J., & Meadows, L. (2002). Standards-based reform and its unintended consequences: Implications for science education within America's urban schools. *Journal of Research in Science Teaching*, 39(2), 114–127.

Shavelson, R. J. (2003). *On The Integration of Formative Assessment in Teaching And Learning with Implications for Teacher Education*. For the Stanford Education Assessment Laboratory and the University

of Hawaii Curriculum Research and Development Group. http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/On%20the%20Integration%20of%20Formative%20Assessment_Teacher%20Ed_Final.doc.

Shavelson, R. J. (2006). On the integration of formative assessment in teaching and learning with implications for teacher education. Paper prepared for the Stanford Education Assessment Laboratory and the University of Hawaii Curriculum Research and Development Group. http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/Paper.htm.

Shavelson, R. J. (Guest Ed.). (2008). *Applied Measurement in Education*, 21(4), 293–403.

Shavelson, R. J., & Ruiz-Primo, M. A. (1999). On the psychometrics of assessing science understanding. In J. J. Mintzes, J. H. Wamhersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view*. New York: Academic Press.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ... Yin, Y. (2008). On the Impact of Curriculum-Embedded Formative Assessment on Learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues Practice*, 28(3), 32–37.

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teacher and Teacher Education*, 20, 163–183.

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.

Sikorski, T. R., & Hammer, D. (2010). A critique of how learning progressions research conceptualizes sophistication and progress. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers*. Chicago, IL: International Society of the Learning Sciences.

Singer-Gabella, M., Cartier, J., Forman, E., Knapp, N., Kannan, P., Shahan, E., ... Barrick, N. (2009). Contextualizing learning progressions for prospective elementary teachers of mathematics and science. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Smith, J., diSessa, A., & Roschelle, J. (1993 /1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115–163.

Smith, C., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth graders' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, 18(3), 349–422.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, 14, 1–98.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology and educational theory and practice*. Albany, NY: State University of New York Press.

Taber, K. S. (2000). Multiple frameworks?: Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education*, 22(4), 399–417.

Tobias, S. (1993). *Overcoming math anxiety*. New York: Norton Company.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Philadelphia: Open University Press.

Tunstall, P., & Gipps, C. (1996). 'How dos your teacher help you make your work better?' Children's understanding of formative assessment. *The Curriculum Journal*, 7(2), 185–203.

Valli, L., Croninger, R. G., Chambliss, M. H., Graeber, A. O., & Buese, D. (2008). *Test-driven: High stakes accountability in elementary schools*. New York: Teachers College Press.

Van Sledright, B. A. (2010). *The Challenge of rethinking history education: On practice, theory, and policies*. New York: Routledge.

van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences*, 6(2), 227–269.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching*, 38(5), 529–552.

Warren, B., & Rosebery, A. S. (1995). “This question is just too, too easy!” Perspectives from the classroom on accountability in science. Santa Cruz, CA: National Center for Research on Cultural Diversity and Second Language Learning.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling and meta-cognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118.

Windschitl, M., Thompson, J., & Braaten, M. (2011). Ambitious pedagogy by novice teachers? Who benefits from tool-supported collaborative inquiry into practice and why. *Teachers College Record*, 113(7), 1311–1360.