ARGUMENT-DRIVEN FORMATIVE ASSESSMENT
FOR CONCEPTUAL SCIENCE LEARNING

Jonathan T. Shemwell
Stanford University School of Education

Erin Marie Furtak
University of Colorado, Boulder

Abstract

Discussion and discussion-centered formative assessment in science classrooms often focus on students’ direct experience of science phenomena. According to the rational view of knowledge construction, students’ can be thought of as reasoning from evidence in discussion to construct explanations. The application of scientific argumentation to science classrooms exemplifies the rational view and suggests that high quality reasoning from evidence and conceptually rich student talk should or at least can be mutually supporting. We investigate the interaction of conceptual talk and reasoning from evidence in videotape transcripts of six middle school teachers conducting evidence-centered classroom discussions intended to foster the development of students’ conceptual knowledge. These discussions, designed to be the same for all six teachers, are explicitly cast as scientific argumentation and take place within a research-developed formative assessment process. We transcribe and code discussions in order to identify whether students’ assertions about a particular concept consist of well-supported arguments (i.e. are supported by evidence) and whether these amount to or conjoin with rich conceptual talk in which students describe phenomena in their own words or advance explanatory mechanisms. We find that well-supported scientific arguments rarely consist of or link to rich conceptual talk, and that most of the rich conceptual talk we observe consists of poorly supported scientific arguments. Using a detailed analysis of sample transcripts, we connect this divergence to design aspects of the formative assessment we study as well as to potential limitations inherent in scientific argumentation as a form of discussion and discussion-based formative assessment for conceptual science learning.
Argument-Driven Formative Assessment

and Conceptual Science Learning

Decades of research have established that student conceptions of the natural world are resistant to change (S. Carey, 1999; S. Carey, 1985; Chi, 2005; Clement, 1982). Teachers explain scientific theories, but students do not believe or understand them, and they cling to their naïve ideas. Learning is more effective when teachers facilitate students’ active construction of knowledge, taking as building blocks those things students already know (National Research Council, 2000). By engaging students in discussion in which they articulate and reflect on their developing conceptions of scientific phenomena, teachers can help students to move from naïve to scientifically accepted ideas (Brown & Campione, 1994; Minstrell & Van Zee, 2003; Van Zee & Minstrell, 1997).

When carefully conducted, discussion engages students in a constructive process of explaining, reflecting upon, and defending their thinking at the critical juncture between new ideas and the prior knowledge needed to make sense of them. Such conceptually rich discussion facilitates science learning because it brings important student thinking into a public, meta-cognitive workspace where ideas can be shaped and refined. Rich discussion supports science teaching for this same reason: to make student thinking visible and explicit, and to provide the impetus and support for refining that thinking, are the essential elements of formative assessment (Sadler, 1989).

Formative assessment, or assessment that facilitates in-progress learning, is the process of eliciting information from students in relation to explicit learning goals, and then taking action in the form of instructional feedback to move students toward those goals (NRC, 2001). Black & Wiliam (1998) have found formative assessment to have a significant positive impact on student
learning, and on narrowing achievement gaps. However, studies indicate that enacting formative assessment presents a significant challenge to science teachers (Furtak et al., 2008).

In science classrooms, discussion and discussion-based formative assessment often center on students’ direct experience of phenomena, as in White & Gunstone’s (1992) Predict-Observe-Explain instructional sequence. However, there are divergent views about how this experience should be characterized and how it should interact with the sense-making process within discussion. Researchers taking a rational view of knowledge construction tend to characterize students’ experience of phenomena as accumulating scientific evidence, and frame students’ sense-making efforts in discussion as deliberations approximating those of scientists (Driver, Newton, & Osborne, 2000; Duschl & Gitomer, 1997; Duschl & Osborne, 2002). According to the rational view, students should reason from evidence during conversations to develop and refine their scientific explanations. Others, especially those primarily concerned with notably difficult science content, argue that changes in students’ views involve a conceptual leap by students which, though motivated and informed by direct experience of phenomena (i.e. scientific evidence), is fundamentally dependent on conceptual structures introduced and sustained in discussion by the teacher, structures that are not necessarily warranted by evidence (Clement, 1993; Minstrell, 1982; Minstrell & Van Zee, 2003). These divergent views suggest very different orientations for classroom discussion.

This paper explores the rational view of knowledge construction within discussion and discussion-based formative assessment, particularly as this view applies to the learning of challenging science concepts. Specifically, we take up the question of how students’ reasoning from empirical evidence within discussion relates to their articulation of emerging conceptions of phenomena, which we call conceptual talk. We study this interaction of evidence and conceptual
talk in a formative assessment context in which classroom discussion is explicitly cast as scientific argumentation, a genre of discourse in which reasoning from evidence is meant to be the primary means by which participants construct understanding.

Background

In this section we first define evidence-based reasoning, and then review some of the ways in which conceptual talk and reasoning from evidence might be expected to interact within the process of scientific argumentation. We contend that scientific argumentation is an important context for studying the relationship of conceptual talk and evidence-based reasoning for two reasons. First, scientific argumentation exemplifies the rational view of knowledge construction within discussion. Second, it has been compellingly argued to be an effective means of conceptual science learning.

Evidence-Based Reasoning

Students’ ability to reason from evidence and participate in scientific argumentation is considered a major objective of science education reform (American Association for the Advancement of Science, 1993; NRC, 1996). According to Duschl and Gitomer (1997), this involves “the development of thinking, reasoning, and problem-solving skills to prepare students to participate in the generation and evaluation of scientific knowledge claims, explanations, models, and experimental designs” (p. 38). Research on discussions in science classrooms have found that episodes of high-level reasoning are rare; in fact, discourse is more often characterized by student claims that are not supported by evidence (e.g. Jimenez-Aleixandre et al. 2000; Osborne, Erduran, & Simon, 2004).

Scientific Argumentation and the Rational View of Knowledge Construction
Defining scientific argumentation. We follow Duschl and Osborne (2002) who in turn draw on Suppe (1998) to define scientific argumentation as “the special case when [a dialogue between two individuals] addresses the coordination of evidence and theory to advance an explanation, a model, a prediction, or an evaluation” (Duschl & Osborne, 2002, p. 55). We attribute three essential properties to scientific argumentation. The foremost is that priority is given to evidence as the basis for asserting the validity of knowledge claims. The second is that argumentation is a social process amounting to an exchange of ideas focused on differing points of view. Third, argumentation has the purpose of building and refining explanations of the natural world.

For the purpose of comparing scientific argumentation to other forms of discussion, the second and third of its three properties are less important. Differences in viewpoint and explaining observed phenomena are common to many forms of science discussion. Thus, the signature feature of scientific argumentation is its clear priority on reasoning from evidence.

What is evidence? In science, and therefore in science classrooms, valid evidence is empirical. It consists of observation statements that have been submitted to the most straightforward testing by the senses as possible, and have passed those tests (Chalmers, 1999). For science classrooms focused on learning science concepts and principles (as opposed, say, to exploring socio-scientific issues), evidence often evolves from shared, immediate experience of phenomena. A critical stance towards evidence is essential to scientific argumentation. Students must be challenged on—and learn to challenge—what “counts” when deliberating what they accept to be true based on evidence.

learning by contending that students, like scientists, are best convinced of (or against) their beliefs by evidence. As Duschl and Osborne state, a student’s construction of scientific knowledge is necessarily the construction of an argument:

The construction of an explanation requires…the construction of an argument that relates models or theories to a body of available evidence. For example, the task of convincing a skeptic that the Earth is a sphere, approximately, or that day and night are caused by a spinning Earth and not by a moving Sun, cannot be done without engaging in a dialogue at the core of which lies the construction of an argument. Rational belief rests on good reasons which consist of data, warrants, and underlying theoretical tenets which acquire epistemic force through critical evaluation against agreed criteria (Duschl & Osborne, 2002, citing Siegel, 1989).

We interpret the above statement to mean that students’ construction of knowledge within phenomena-centered classroom discussion should fundamentally be a rational process of change in belief warranted by evidence. These authors also point out, however, that while they expect reasoning from evidence can change students’ thinking about science concepts, empirical studies showing how this happens, and to what extent, and under what limiting conditions, is almost nonexistent. They urgently call for more research to bring these issues to light.

*Argument-driven formative assessment.* Duschl and Gitomer’s (1997) assessment conversation exemplifies a type of formative assessment focused on advancing students’ scientific understanding through the interaction of conceptual talk and reasoning from evidence. The assessment conversation is an instructional dialogue specially designed to embed assessment into classroom discussion. These conversations engage students in discussion of diverse student-generated ideas focused upon evidence and scientific ways of knowing. The formative
assessment loop functions when students make their thinking visible by articulating their emerging conceptions while teachers guide the discussion to augment or refine those conceptions. Thus, the assessment conversation can be thought of as an argument-driven formative assessment process in which, with the teacher as moderator, students argue their differing ideas based on the evidence that they can muster for and against them. Again, knowledge construction is expected to be a rational process by which consideration of evidence facilitates the construction and refinement of scientific explanations.

Against the Rational View of Learning in Discussion

The need to work with and not against prior knowledge. The idea that students should reason from evidence to construct knowledge is addressed in the body of work on conceptual change. Research in this field, having pointed up both the intransigence and incoherence of naïve student conceptions, has converged on the view that it is necessary to work with, rather than against, students’ initial ideas (diSessa, 2006; NRC, 2007). From this perspective, moving from a misconception to a more scientific understanding involves refining, rather than replacing, student’s prior knowledge. The theory that students learn by constructing rational arguments backed up by evidence—the theory of learning by argumentation—would seem to favor the replacement as opposed to the refinement of ideas, because it fails to address how evidence is meant to interact with naïve conceptions. While proponents of the rational model of conceptual change have attempted to accommodate refinement as opposed to replacement (Strike & Posner, 1992), there is no empirical basis for this model in either classroom or individual learning.

In one especially prominent example, Minstrell (1982) and Clement (1993) show that reasoning from evidence may not embrace all the forms of thinking that might be needed to bring about conceptual change. These authors describe the salutary effect of engaging students in
discussion of analogies between certain difficult scientific conceptions and more readily apparent “anchoring” conceptions. In the example they present, students can more easily “see” that an inert table can exert a force on a book if they compare the condition of the table contacting the book to the action of a spring pressing on the book. Clement points out that in designing bridging analogies, priority is given to relating new ideas to things students already know as opposed to making them scientifically defensible, so that a student’s crossing of the analogical bridge falls well short of reasoning from evidence.

The need for selective emphasis in classroom discussion. Tabak and Reiser (1999) point out another potential source of divergence between reasoning from evidence and the construction of conceptual understanding within science discussion. These researchers depict the need for students and teachers to be selective about what they emphasize in students’ explanations. To do this, they illustrate one teacher’s successful effort to help students in a guided inquiry classroom string together isolated statements of facts into coherent explanations. They point out that the teacher sets aside strict requirements for students to reason from evidence to enable a preferential focus on constructing more elaborate explanations. They argue that in this teacher’s case, this setting aside of evidence-based reasoning is necessitated by the many challenges that her students face as they struggle to articulate the relationships between causal conditions. In other words, they contend that teachers and students have limited bandwidth for reasoning in discussion. The implications for scientific argumentation are clear: to establish reasoning from evidence as a rigid requirement up front is to run the risk of limiting students’ individual and collective capacity for conceptual reasoning.
Research Questions

According to the rational view of knowledge construction as we have presented it, there should be some relationship between students’ use of evidence and the explicitness of their conceptual talk when they are speaking about challenging science phenomena in discussion. Our purpose was to look for this relationship in the argument-driven formative assessment context we studied. We therefore asked two questions of our data:

1. With what frequency did students’ high quality reasoning from evidence and explicit conceptual talk tend to conjoin or disjoin in discussion?

2. Within observed conjunctions and disjunctions, what conditions tended to prevail?

We addressed the first research question by coding videotaped transcripts of student dialogue that focused on a particular concept, sorting dialogue into various categories according to the quality of students’ reasoning from evidence and the explicitness of their conceptual talk. We took up the second question using detailed explications of example transcripts from the various coding categories.
Context

In this section, we describe the larger research project on formative assessment from which data for this study was drawn, including a brief overview of the curriculum in which it was embedded, as well as the conceptual domain in which it was based. We give particular attention to those aspects of the formative assessment design intended to promote scientific argumentation that would also be conceptually rich discussion.

Curriculum

The data we analyzed were collected as part of the Romance Project, an NSF-funded small-scale randomized trial investigating the impact of embedded formative assessment on student learning (Shavelson et al., 2008; Yin, Tomita, & Shavelson, 2008 both citations provide overviews of study and findings). The unit explored in the Romance Project was taken from *Foundational Approaches to Science Teaching* [FAST] (Pottenger & Young, 1992), a standards-aligned middle school science curriculum. The investigations in FAST combine inquiry-based activities with a conceptual progression of student understanding built through a sequence of inquiry tasks. The unit studied in this project, *Introduction to the Properties of Matter*, consists of 12 investigations designed to support students’ construction of the concept of relative density as it relates to sinking and floating. FAST models scientific practices for students, incorporating multiple roles for student and teacher. While students form hypotheses, perform experiments, analyze data, and develop consensus on scientific ideas, the teacher acts as director of research and colleague who stimulates and makes possible deeper inquiry (Pottenger & Young, 1996).

Embedded Formative Assessments

FAST moves students through investigations that explore the effect of mass on sinking when volume is controlled, then volume when mass is controlled, density, and finally the relative
density of an object and its surrounding fluid. The Romance Project modified FAST by adding a series of formative assessment “suites” embedded at several different points in the fast curriculum (for more information on the development and implementation of these assessments, see Ayala et al., 2008 and Furtak et al., 2008). Each formative assessment suite lasted about three class periods and combined a set of evidence-rich activities with scientific argumentation, including the interpretation of a graph, a Predict-Observe-Explain (POE) activity (White & Gunstone, 1992), and a short writing prompt asking students to compose their own general explanation for why things tend to sink or float.

To simplify the analysis contained in this paper, we examined data from a single formative assessment suite, one in which students were learning how and why differences in volumetric displacement affect objects’ tendency to sink or float. These ideas are explored in the seventh investigation of the FAST curriculum, in which students look for relationships between displaced volume, total volume, and the sinking and floating of objects when mass is held constant. While density is not explicitly introduced in either the seventh investigation of FAST unit or the formative assessment unit we studied, the idea of density is salient in terms of changes in the relationship of mass and volume across multiple objects. A summary of the formative assessment prompts embedded in the investigation we explored is shown in Table 1.

Table 1. Description of the Four Formative Assessment Prompts

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description: Students were asked to...</th>
<th>Purpose: To determine if students could...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph</td>
<td>Evaluate, correct, and interpret a</td>
<td>Correctly identify errors in a graph</td>
</tr>
</tbody>
</table>

- 11 -
<table>
<thead>
<tr>
<th>Interpretation</th>
<th>graph of mass versus displaced volume of objects that sinking or float in water.</th>
<th>and to self-assess their work (Sadler, 1989).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict-Observe-Explain</td>
<td>- Given the mass and displaced volume of three sealed, opaque, plastic bottles of equal mass but differing volumes, predict whether each will sink or float in water.</td>
<td>- Interpret quantitative evidence from the graph to explain sinking and floating.</td>
</tr>
<tr>
<td>Constructed Response</td>
<td>- Respond to an open-ended question that was central for the unit: Why do things sink and float?</td>
<td>- Transfer knowledge about sinking and floating from the investigation to the new experimental setting, and provide accurate reasoning for the outcome (White &amp; Gunstone, 1992).</td>
</tr>
<tr>
<td>Predict-Observe</td>
<td>- Given the mass and total volume of a fourth bottle, predict and observe whether it will sink or float in water.</td>
<td>- Provide an explanation for sinking and floating with supporting evidence and/or examples</td>
</tr>
<tr>
<td>Constructed Response</td>
<td>- Extend existing knowledge to an experimental setting with cognitive demands beyond the present level.</td>
<td></td>
</tr>
</tbody>
</table>

Once students’ initial ideas were elicited in the formative assessment process, scientific argumentation was to be the primary means of advancing students in their understanding (Figure 1). It was therefore the formative assessment designers’ explicit goal that students’ construction of well-supported scientific arguments amount to or precipitate conceptually rich discussion of the pertinent science phenomena. Evidence to fuel the argumentation process could come from
either the formative assessment activities themselves or from the many hands-on experiences in the main curriculum. Teachers were expected to begin by eliciting students’ initial ideas and attempted explanations, together with their reasons for these, in written prompts (Furtak & Ruiz-Primo, 2008). They were then to cluster student responses into rough categories before reading or posting them on the board with students’ names removed. Students, guided by the teacher, were to engage in whole-class scientific argumentation to debate the merits of differing ideas and explanations with respect to empirical evidence, to refine those ideas and explanations, to advance new ones, and so forth.

![Diagram of the intended placement of argumentation in the formative assessment process](image)

**Figure 1.** Intended placement of argumentation in the formative assessment process (Stanford Education Assessment Laboratory, 2003)

**Method**

**Participants**

The six teachers implementing the formative assessment represented various backgrounds and levels of experience. In two of the classrooms, the curriculum was implemented in 6th grade, and in the remaining four, the 7th grade. Class sizes ranged from 20 to
31, and class-period length also varied. School contexts ranged from those serving mostly white suburban middle class children to those serving mostly minority, lower SES populations. See Shavelson et al. (2008) for a more complete account of participant characteristics.

Sources of Data

Videotapes and transcripts of the six teachers implementing the seventh investigation in the FAST unit and its accompanying formative assessments comprised data for this study. Teachers took between one to three class sessions to complete the formative assessments on the effect of volumetric displacement on density and buoyancy. Working from the transcripts and the videotapes, we identified the whole class discussions (argumentation sessions) for each teacher and time-marked these at one-minute intervals (196 minutes across all teachers).

For each minute of transcript, we then determined whether or not any student asserted a causal linkage between volumetric displacement and buoyancy (causal volume claim). A student assertion qualified as a causal volume claim if he or she mentioned volume (or corresponding terms such as displacement), or a property of volume such as size, or a consequence of volume, such as a less intense distribution of mass, as a cause or condition of sinking and floating. Working independently with a greater than 20% random sample of data, we achieved 97% inter-rater agreement identifying students’ causal volume claims. A single researcher then identified causal claims in the remainder of the data independently. We separated a total of 75 instances for further analysis.

Coding Procedure

In this section we describe how we segmented transcripts into units of analysis incorporating student-teacher and student-student interactions around volume claims, and how
we coded the dialogue along the two principle dimensions of our analysis: conceptual explicitness and reasoning from evidence.

*Segmenting data into reasoning units.* We segmented the data around each volume claim to form *reasoning units*—continuous strings of student and teacher conversational turns focused on a given volume-related assertion or set of related assertions. Reasoning units corresponded roughly to topics of conversation in the sense that they encompassed all of the student utterances explicitly relating to a given assertion (Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, 2008). Each reasoning unit started with the teacher prompt leading to the causal claim and ended when the claim or related claims and qualifiers were dropped from the conversation. We completed this segmenting step independently, achieving at least 80% agreement on a greater than 20% random sample of causal volume claims. Having established this level of reliability, we independently segmented the remainder of the data and adjudicated any differences. Since some causal volume claims were related by dialogue and therefore combined within a single reasoning unit, we ended up with a slightly lower number of reasoning units than causal volume claims (65 reasoning units vs. 75 causal volume claims). Reasoning units varied in length from less than 20 seconds to 6 minutes.

*Coding reasoning units.* Each reasoning unit was coded at two levels of conceptual talk and three levels of reasoning from evidence. Each set of codes will be explained in detail below; codes are summarized in Table 2. Here we introduce the convention that key utterances relevant to conceptual explicitness are underlined, while those related to use of evidence are in bold type.

Table 2. Description and Examples of Conceptual Explicitness Code

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Example Student Statements</th>
</tr>
</thead>
</table>

- 15 -
**Conceptual Talk**

**Conceptually Implicit**
- References to displaced volume, its contributing factors, or its effects within the stated claim are restricted to terms pre-sanctioned by the curriculum, class or teacher: *volume, displaced volume, displaced water, total volume, displacement.*
- Causal reasoning chain describes no mechanism for the affect of volume on buoyancy.

“[The bottle sank] because the amount of water displaced was less than the mass.”

**Conceptually Explicit**
- References to volume are stated in students’ own words (underlined).
- Student’s causal reasoning chain contains a stated or described intervening condition in the causal reasoning chain between volume and buoyancy (underlined).

- “Because the smaller one, it wasn’t as large, it took a little more in it. It’s going to go through the top more and if you put it in the big one, it’s going to have space.”
- “It’s [the larger bottle is] going to float because there’s a lot of air in it [the relationship between air and larger volume is inferred].”
- “When you have more volume, it looks like the mass spreads out...”
### Quality of Reasoning

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Sample Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim-Based</td>
<td>Statement of what something will do in the future (prediction), or is happening in the present or past (conclusion or outcome)</td>
<td>“…if there’s less amount of mass, or the same amount of mass as the volume, it will float.”</td>
</tr>
<tr>
<td>Data-Based</td>
<td>Claim backed up by a single observable property.</td>
<td>“The bigger one could subsurface, and that one could float because the big one, if it went to the middle, then it would displace like the same amount because it's bigger than the middle one, and the middle one could float…because of the size difference.”</td>
</tr>
<tr>
<td>Evidence-Based</td>
<td>Claim supported or backed up by statements describing a contextualized relationship between two observable properties or a contextualized relationship between a property and an observable consequence of that property</td>
<td>“It’s going to float… because the displaced volume equals 70 and the mass of the bottle is 70.”</td>
</tr>
</tbody>
</table>

*Teacher Support*
Evidence provided by teacher
Teacher provides elements of reasoning (data, or evidence)
T: Remember our cartons? The milk cartons?
S: Yeah.
T: Where we had the same mass…but different size cartons and they displaced…they sank to different levels.

Evidence provided by student
Student provides elements of reasoning (data, or evidence)
S: Because it was the same amount of mass in a smaller…in a smaller, area or, container.

*Conceptual talk.* For conceptual talk, only student statements relating to volumetric displacement were coded. This “volume talk” was coded as either conceptually explicit or conceptually implicit. The higher level (conceptually explicit) code was assigned if any student’s claim or supporting statements within the reasoning unit met either of two criteria. The first of these was for a student to describe volumetric displacement or its characteristics or consequences in his or her own words (i.e. in words other than the canonical terms displaced volume, volume, total volume, displaced water, or displacement). The alternate criterion required a student to attempt to state or describe some manner of intervening condition relating volume as a cause to sinking or floating as an effect. This intervening condition did not have to be scientifically acceptable.

*Quality of Reasoning.* In contrast to conceptual talk, we coded the quality of reasoning from evidence for both teachers and students (at least initially), and we looked for this reasoning
anywhere in the dialogue segment (reasoning unit). For this coding, we used a modified version of the “Using Evidence” analysis framework (Furtak et al., 2008), differentiating between the kinds of reasons (a.k.a. support/backing) participants gave for their ideas. The codes include reasoning units in which claims have no support (claim-based), are based only on a single variable or observation (data-based), or on the relationship between variables or properties (evidence-based).

Teacher support. Finally, to determine whether it was students or teachers who provided evidentiary support (i.e. data and evidence) for claims, we coded each of the 40 reasoning units in which data or evidence was used as support for claims to determine with whom this support originated (teacher or student). If both students’ and teachers gave support for their claims, then the reasoning unit was coded as student-provided at the level applicable to the student utterance. Since our focus was solely on student reasoning and talk, we re-designated those reasoning units in which teachers were the sole providers of data or evidence as claim-based student reasoning. There were 7 such reasoning units in all. Table 3 illustrates one of them.

Inter-rater agreement. We obtained 80% or more inter-rater reliability on a greater than 20% random sample for each of these three codes. Once reliability was established, we independently coded the remainder of the data and adjudicated any differences. A sample coded segment is shown in Table 3 to illustrate the manner in which we applied the codes to each reasoning unit. Table 3 is coded as claim-based student reasoning since evidence introduced in the discussion comes from the teacher. Meanwhile, it is coded as conceptually explicit, because the student states a mechanism, trapped air, as an intervening condition connecting the volume (size) of the object and its tendency to sink or float.
Table 3. Example Coded Transcript.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher:</td>
<td>Why? Why did number one sink? I'm interested in somebody who got it wrong and had to rethink it. Or Alyssa who wrote, “because I saw it.” And so she had to make up a reason, too. So, Alyssa, do you want to start off and give us a reason why it sank?</td>
<td>Conceptually Explicit Student Talk: least amount of trapped air in the container?</td>
</tr>
<tr>
<td>Alyssa:</td>
<td>Because it had…the least amount of…trapped air in the container?</td>
<td></td>
</tr>
<tr>
<td>Teacher:</td>
<td>Trapped air in the container. Okay. Most of our investigations depend on trapped air?</td>
<td>Claim-based Reasoning: Student has no knowledge of air or empty space in the sealed container.</td>
</tr>
<tr>
<td>Alyssa:</td>
<td>Not most of them.</td>
<td></td>
</tr>
<tr>
<td>Teacher:</td>
<td>Going back to our liquids and vials, right, we have one that has more trapped air. [Teacher holds up the graduated cylinder in which different vials are floating</td>
<td></td>
</tr>
</tbody>
</table>
at different depths.] Right?

Student:   Whoa.

Teacher:   This guy [vial floating lower] has about a third of it is trapped air. This guy [vial floating higher] has like about half of that – about a sixth trapped air. Yet this one…is above this one. You think it’s trapped air? If it’s trapped air, please explain these two.

---

**Data Analysis**

Having identified, segmented, and coded reasoning units as described above, we sorted the 65 reasoning units into a 2 x 3 (conceptual explicitness x reasoning from evidence) matrix. In a secondary analysis we then developed descriptive characterizations of the type of talk and reasoning that prevailed in each cell of the matrix.

**Results**

The results of our data analysis indicate that conceptually explicit talk and high-quality reasoning from evidence did not go hand in hand in the formative assessment process. We present our findings in three sections. In the first we examine the frequency with which various levels of reasoning conjoined or disjoined with conceptual explicitness. In the second, we present examples of student dialogue to illustrate the manner in which these conjunctions and disjunctions tended to occur. In the third, we present a detailed explication of two of these transcripts as they relate the rational view of knowledge construction in discussion, including argument-driven formative assessment.
Overview of Coding Results

About half of the 65 reasoning units consisted of claim-based reasoning (32 reasoning units, 49%), with the second-highest group being evidence-based reasoning (24 reasoning units, 37%), and the smallest group making up data-based reasoning (9 reasoning units, 14%) We did not see noteworthy differences among teachers across these categories. In terms of conceptual explicitness, 24 reasoning units (37%) were coded as conceptually explicit and 41 (63%) were conceptually implicit. Again, we saw a fairly uniform effect across teachers. We present these results in Table 4.

Table 4. Frequency of Reasoning from Evidence and Conceptual Talk

<table>
<thead>
<tr>
<th>Quality of Reasoning</th>
<th>Totals</th>
<th>Claim-Based</th>
<th>Data-Based</th>
<th>Evidence-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual Talk</td>
<td>Implicit</td>
<td>41</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Explicit</td>
<td>24</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

Reasoning units featuring evidence-based arguments involving conceptually explicit talk were rare (8 reasoning units, 12%), whereas those incorporating lower level (claim- or data-based) reasoning with conceptually implicit talk were more frequent (25 reasoning units, 38%). More importantly, there was a marked divergence between reasoning from evidence and conceptual talk. First, for those segments featuring evidence-supported claims, twice as many were conceptually implicit as conceptually explicit (16 vs. 8 reasoning units). Second, students made many conceptually explicit statements that only amounted to claim- or data-based reasoning (16 reasoning units, 25% overall).
From the standpoint of the conceptual sense-making which the formative assessment process was intended to promote, much of the more explicit conceptual talk did not amount to well-reasoned scientific arguments, and most instances of evidence-based reasoning did not amount to conceptually explicit talk. In short, better use of evidence and higher level conceptual talk did not go together.

**Describing Coded Categories of Reasoning Units**

In this section we present coded and explicated examples of the various categories of reasoning unit we obtained in the coding process, together with descriptions of the prevailing character of student-teacher interactions within those categories.

_Collapsing claim and data-based reasoning units._ We found the 9 reasoning units coded as data-based reasoning to feature extremely limited evidentiary support for student’s claims and attempted explanations. Consequently, we collapse claim-based and data-based reasoning units in this section. This collapsing results in four types of reasoning units depending on whether or not talk was conceptually explicit and whether or not students successfully reasoned from evidence. Table 5 gives a summary description of each of these four types. We explicate each type in separate transcript excerpts (Tables 6-9), then discuss what these explications mean for the relationship between conceptual talk and reasoning from evidence within discussion and discussion-based formative assessment.
Table 5. Describing Four Categories of Reasoning and Talk

<table>
<thead>
<tr>
<th>Prevailing Reasoning From Evidence</th>
<th>(Data or Claim-Based)</th>
<th>(Evidence-Based)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implicit</strong></td>
<td>If/then decision rule comparing numerical values for mass and volume to determine whether an object will sink or float. <em>If an object’s mass is more than its volume, it’ll sink.</em></td>
<td>Same if/then rule as at left, but with an added “because” statement in which numerical values were used to assert the truth of the rule</td>
</tr>
<tr>
<td><strong>Explicit</strong></td>
<td>- Described supposed features of classroom artifacts (such as trapped air or mass spread out/compacted together) as reasons why things might sink or float.</td>
<td>- Same as at left, but also made claims about sinking or floating based on non-numerical evidence in the form of relative magnitudes of mass or volume</td>
</tr>
<tr>
<td></td>
<td>- Supplied examples of higher or lower density sinking or floating items (such as a sponge) from outside the common experience of the classroom.</td>
<td>- Supplied examples of lower or higher density sinking or floating items such as wood or rocks that were within the classroom’s evidence pool.</td>
</tr>
</tbody>
</table>
**Conceptually implicit talk with evidence-based reasoning.** Reasoning units coded as conceptually implicit talk with high level reasoning from evidence invariably consisted of straightforward decision rules comparing objects’ mass and volume. The rule would generally involve some variation on the statement, “If the object’s mass is more than its volume, then it sinks; if not, it floats.” To provide evidentiary support for claims invoking this rule, students’ generally cited numerical figures for an object’s (or a group of objects’) mass in grams and volume in milliliters. Table 6 provides an example that illustrates this occurrence, in which a teacher asks students to raise their hands and then counts the number of students who believe one of the bottles is going to float. Following the convention introduced in the methods section, key utterances relevant to the reasoning from evidence coding are in bold type, while those relating to conceptual explicitness are underlined.

Table 6. Transcript Excerpt Illustrating Conceptually Implicit Talk with Evidence-based Reasoning

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue</th>
<th>Coding Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teacher:</strong></td>
<td>Who thinks it's going to float? One, two, three, four, five, six, seven</td>
<td>Conceptually Implicit: Sylvia’s expression for displacement, displaced water, is on the list (italics on Table 2) of canonical terms which do not qualify as her own words.</td>
</tr>
<tr>
<td></td>
<td>[Teacher counts raised hands] Why do you think that, Sylvia?</td>
<td></td>
</tr>
<tr>
<td><strong>Sylvia:</strong></td>
<td>Um, because it's got, um, it’s got the displaced water of 70 and it’s also got the mass of 70, around there.</td>
<td>No mechanism is inserted in the causal chain between mass = volume and floating.</td>
</tr>
</tbody>
</table>
Teacher: So because, okay, because the displaced volume equals 70 and the mass of the bottle is 70.

Evidence-based reasoning:
Sylvia provides evidence for her prediction in the form of contextualized relationship between two properties, the mass and displaced water of the bottles.

Sylvia’s prediction amounts to a well-reasoned scientific argument in the sense that it is supported by evidence, but she makes no attempt to articulate why this mass vs. volume relationship might result in floating (strictly speaking, neutral buoyancy); nor does she address what the quantities of 70 and 70 actually mean.

*Conceptually implicit talk with claim- or data-based reasoning.* Claim- or data-based, conceptually inexplicit reasoning units almost uniformly featured the same kind of rule statement that Sylvia makes in Table 7, but in these cases students offered no evidence to back up the rule. Table 7 illustrates this case.

Table 7. Transcript Excerpt Illustrating Conceptually Implicit Talk with Claim- Or Data-Based Reasoning

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue</th>
<th>Coding Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher:</td>
<td>So, Mark, what was your conception of why things sink and float?</td>
<td>Conceptually implicit: Mark’s expression for displacement, water displaced, is on the list (underlined on Table 2) of canonical terms which do not qualify as his own words.</td>
</tr>
<tr>
<td>Mark:</td>
<td>I put . . . if things, objects have the</td>
<td></td>
</tr>
</tbody>
</table>
same mass and volume of water displaced, then they float. If they don't, then they’ll sink. No mechanism is inserted in the causal chain between mass = volume and floating.

[tapping sound as teacher writes on chalkboard]

Mark provides no evidence to support this claim.

Mark: Then they'll float. And if they don't, then they'll sink.

Teacher: Okay. Claudia?

Of note in Table 7 above is the fact that Mark makes an implied reference to empirical evidence in the form of numerical values of the mass and displaced volume of the collection of “things” or “objects” tested in class (line 2). We know that this evidence consists of numbers because there is no way to compare mass and volume in terms of relative magnitudes.

Conceptually explicit talk with evidence-based reasoning. Within the 24 reasoning units coded as conceptually explicit, students either described volumetric displacement in their own words, or they stated some form of volume-related intervening condition that related volume, or mass and volume together, to sinking or floating. Expressions for displacement included such phrases as “the space something takes up” or “the area,” or (in a few cases) simply “bigger,” where, as “bigger” is not just an identifier for a larger bottle but a description of it. Intervening conditions included described conceptions such as statements that things sink when the mass is
tightly packed and float when the mass is more spread out, or examples comparing large, low
density items to smaller, high density ones (e.g. block of wood vs. rock). The most common
mechanism cited by students for how volume affects floating was the presence of trapped air, a
classic naïve conception of buoyancy which we contend can both contribute to more
sophisticated understanding (as a factor in average density) and hold students back in their
thinking as they may over-generalize the effects of air. In Table 8, we see a student, Melissa,
articulate her conceptions of volume and density in terms of “area” and “what’s in the inside,”
“spread out,” and “air pockets.”

Table 8. Transcript Excerpt Illustrating Conceptually Explicit Talk with Evidence-based
Reasoning

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue</th>
<th>Coding Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher: Melissa, you said number one was going to float. Why now do you think it actually sank? Any reason at all you can think about? What did you write on your sheet on the second page for why it sank?</td>
<td>Conceptually Explicit: Melissa expresses displacement in her own words with area, bigger, smaller, like on the inside, etc. She also expresses the effect of increasing volumetric displacement on density and on floating in her own words as spread out. Finally, she invokes the preconception air pockets.</td>
<td></td>
</tr>
<tr>
<td>Melissa: I think maybe it was because, like, the area that it was in. The area of – Teacher: What do you mean by area? I’m not sure what – Melissa: Area is that it was spread out in. So it was in a bigger container, or a smaller container –</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Teacher: Okay, is that area? Is that area? Area, we said was going to be the amount of this surface. That would be area. But if we were going to measure how much space something took up, that would be?

Melissa: Like on the inside.

Teacher: That would be volume.

Melissa: Yeah, volume.

Teacher: Okay, so the volume – say it again,

Melissa: The volume.

Teacher: No, I mean what you were –

Melissa: The volume was lower.

Teacher: The area that the mass was in.

Melissa: Like it was smaller than all the other containers.

Teacher: Okay.

Melissa: So I thought it would float.

Teacher: Yeah, but no. But it sank. So now we’re asking why you think that it actually did sink.

Melissa: Oh, why I think that it did…So now I
think that’s why.

Teacher: Okay, because of volume. Evidence-based Reasoning: Melissa cites

Melissa: Because it was the same amount of mass in a smaller…in a smaller, area or, container.

Teacher: Smaller container. Smaller. Same amount of mass.

Melissa: And I think it wasn’t as spread out.

Teacher: Okay, in a smaller container.

Melissa: Um Hmm. It wasn’t as spread out. Unless there were air pockets in there it
would never [indistinguishable two syllables which sound like “add all”].

In addition to the fact that Melissa makes many conceptually explicit statements
(underlined) in Table 8 she also constructs an evidence-based argument by comparing the mass
and the volume of the bottles and relating these to sinking. The final 5 lines show the clearest
nexus between conceptual explicitness and use of evidence, as the evidence itself is expressed in
a conceptually explicit way (smaller area or container) and linked to a described mechanism for
the effect of increasing volume (spread out). We also see that, in contrast to Sylvia’s use of
evidence in the conceptually lean strong argument of Table 8, Melissa does not refer to
numerical figures for mass and volume, but to relative magnitudes. This use of non-numerical
forms of evidence was a characteristic of all the reasoning units in the conceptually explicit, evidence-supported category.

*Conceptually explicit talk with claim- or data-based reasoning.* Overall, we found that conceptually explicit dialogue amounting to claim- or data-based reasoning from evidence incorporated the same kinds of expressions and descriptions that students offered in the evidence-supported category, including the conceptual model that mass spreads out with increased volume and trapped air as a named mechanism. However, in contrast to arguments such as Melissa’s in Table 8, students did not explicitly relate expressions for volume or described models and mechanisms for density and buoyancy to the pool of empirical evidence available to them. This is not to say that students’ claims were made out of context. All 16 reasoning units in this category featured statements that drew in some way upon relevant classroom experiences of phenomena or on personal experience. However none of these claims and statements was supported by evidence amounting to valid observation statements. Table 9, an instance of data-based reasoning, illustrates this kind of situation:

Table 9. Transcript Excerpt Illustrating Conceptually Explicit Talk with Claim- or Data-based Reasoning

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialogue</th>
<th>Coding Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher: Even though we know that all of the bottles were 70 grams when we massed them out, why do you think that this one would sink and this one would float? What’s the difference here? Take a look at the bottles.</td>
<td>Conceptually explicit: Maxine expresses volume as smaller. Russell names the mechanism more air to relate volume and floating.</td>
<td></td>
</tr>
<tr>
<td>Maxine: Um…I think that one’s smaller than this.</td>
<td>Claim- and data-based reasoning:</td>
<td></td>
</tr>
</tbody>
</table>
Teacher: Yeah, this one’s smaller and this one’s larger [Referring to bottles, one held in each hand]. Therefore, why might this be happening? Russell?

Russell: Could there be more air in this, if they, if they had the same amount of liquid, that one would have more air than that one.

Teacher: Ok, so you’re saying liquids in these?

Russell: Well, well if there was liquids in them then, then, they would probably have to have the same amount of liquid because they were all 70 grams, plus, but, but bottle 2 has more air, more air than bottle

Maxine’s claim is data-supported because she refers to a single observable property, smaller.

Russell’s claim is data-supported because he refers to the single observable property, mass and attempts to relate this to something he has not observed, the presence of air. Russell’s claim that there “could be more air in this,” though a reasonable inference, is not based on evidence which “counts” in the scientific sense. The presence of air is not a valid observation statement because Russell has not looked inside the bottles.

Discussion of Explicated Transcripts

The examples we have presented here, together with the coded categories they represent, suggest two issues likely to be of general concern when designing discussion-based formative assessment learning environments. The first of these concerns the inclusion of numerical values
in the pool of evidence/experience upon which student reasoning is expected to draw. The second issue involves constraints on discussion that might follow from giving priority to reasoning from evidence, as is the case with scientific argumentation.

**Numerical Forms of Evidence and Conceptual Talk**

The close relationship between numerical evidence and lower level conceptual talk is illustrated by Table 6, in which the student Sylvia makes an evidence-supported claim comparing the mass and volume of one of the bottles that amounts to very limited conceptual talk. For this case and others similarly coded, the provision of numerical values for mass and volume facilitates a degree of abstraction away from the concepts the quantities represent. We find evidence of this abstraction when both Sylvia and her teacher are willing to compare the numbers without going into what they stand for, noting in particular the absence of named units in both teacher’s and students’ statements. Moreover, neither Sylvia nor the teacher grapples with what it might mean for the mass and the volume to be “equal.” This abstraction turns out to be helpful in propounding the important “compare mass to volume” decision rule because pure numbers provide a content-free representation that facilitates the direct comparison of the two conceptually different things (see Schwartz, Martin, & Pfaffman, 2005 for an elaboration on this theme). However, this same abstraction, analytically powerful as it may be, readily permits Sylvia and many others to bypass the active sense-making that was the primary goal of the argumentation process. This bypassing of conceptual talk is especially significant when we note that nearly all of the conceptually implicit reasoning units, evidence-supported or not, were concerned with the same “compare mass to volume” rule. The inclusion of numerical data not only short-circuited Sylvia’s conceptual reasoning, it nearly short-circuited the entire formative assessment process.
We conclude from Sylvia’s example, together with many similar examples across all the classrooms in our data set, that the abstraction of physical quantities as numbers can inhibit rich conceptual talk. Those who would set up reasoning from evidence as a means of conceptual learning in discussion should be cautious when including potentially abstract numerical quantities in the evidence pool to be discussed.

*Scientific Argumentation and Conceptual Talk*

The second of our two points is likewise a cautionary one, regarding the use of scientific argumentation as a mode of discussion for conceptual science learning. Here we will draw on Table 9, in which the student Russell imagines air to be the agent of buoyancy for the sealed, opaque aspirin bottles. Our analysis here will be an extended one, as there are several ways to look at Russell’s story.

*Imaginative and therefore untouchable reasoning.* It is our contention that Russell’s and many other students’ assertions of air as a mechanism for floating are necessarily unsupported claims because the evidence pool available to them does not address the presence of air or any buoyant effects it might have. Nevertheless, from the perspective of conceptual talk, Russell’s and other students’ invocation of air is indeed valuable, as when Russell relates the amount of air to the amount of liquid in the bottles to attempt a conceptually coherent model for differences in density. Rather than evidence, suggestive features of the bottles provide the necessary support, or at least cue, for Russell to use his own words to express what is the same about the bottles (the amount of matter), and what is different (the amount of empty space), in terms of things that he readily understands (air and liquid). When it comes to productive discussion, the validity of Russell’s evidence—the truth of what’s inside the bottles—is not important; it’s what Russell imagines to be inside that propels his thinking. However, to sustain and develop Russell’s
trapped air claim would necessarily be a problem for Russell’s teacher (or any teacher) so long as the discussion is explicitly cast as scientific argumentation. This is because, if the teacher is to faithfully represent scientific ways of knowing, she must discriminate against explanations which are not backed up by evidence.

A necessarily rigorous approach to what counts as evidence. Once could argue that, in relegating Russell’s “could there be more air in this?” from Table 9 to the category of unsupported claim, we are putting too fine a point on the distinction between evidence and non-evidence. The presence of air in the relatively lightweight containers is a reasonable inference, and Russell has probably based it on a wealth of direct, personal experience with similar sealed, lightweight containers. Indeed, Russell is right: there is air inside the bottles. But scientific evidence is valid to the extent to which it has been submitted to and passed the most straightforward and sensible tests possible, such as opening the bottles to see what is inside them. This distinction is crucial because engaging in argumentation must teach students to think critically about the validity of evidence as the basis for “what we know.” Indeed, we might imagine a wise (and well equipped!) teacher revealing the faultiness of Russell’s and others’ assumptions by cutting open the bottles to show them to be filled with a low density solid material.

A good hypothesis but a poor explanation. We must also acknowledge the hypothetical nature of Russell’s reasoning and its genuine scientific value in this respect. As a hypothesis, his trapped air claim has merit because it is falsifiable. In the event that bottles were to be filled with the low-density solid material, so that the teacher could cut them open to prove Russell wrong, this falsification could engender further hypotheses and further advancement in understanding (see Lawson, 2003 for additional development of this hypothetico-deductive view
of argumentation). Nevertheless, scientific argumentation is fundamentally concerned with what can be said based on the evidence, not on assumptions and guesses. Furthermore, in the context we present here, explanations, not hypotheses, are the goal of the activity. Therefore, Russell’s claim, as potentially productive as it is, cannot be a well-reasoned scientific argument.

_How much evidence is enough?_ Although, from the standpoint of scientific argumentation, Russell’s and other students’ claims that air causes things to float cannot be entertained in the situation as we found it to be, one can argue that such claims can readily be incorporated in argumentation pedagogy and could have been for these students. All that would have been required would be for the teacher to have set the table, so to speak, with additional evidence addressing the students’ conception. This evidence might have taken the form, for instance, of observations of the buoyancy of objects of equal mass and volume which contain varying amounts of observable trapped air. Nonetheless, our objection to the potentially constraining effect of scientific argumentation still stands. There will inevitably be points in discussion in which student thinking, particularly thinking that is rooted in prior experience, is not addressed by evidence at hand, and that, lacking this evidence, it will be problematic to nurture and develop ideas in a way that does justice to what scientific argumentation is supposed to be. In short, to restrict classroom discussion to scientific argumentation is to run the risk of pushing valuable student ideas aside.

**General Discussion**

In summary, we have examined two things expected to go together in scientific argumentation and found them to be divergent, at least in the formative assessment process we studied. We evaluated the richness of students’ conceptual talk in terms of how explicitly students articulated the meaning of and relations between concepts they used in their verbal
assertions. We rated the quality of students’ reasoning from evidence on the extent to which students supported their claims with empirically valid observations, as they were called to do in the process of scientific argumentation in which they were involved. Examining these two variables across six classrooms, we found that students did successfully reason from evidence to construct scientifically valid arguments, and they did engage in rich conceptual talk. However, a positive relationship between rich talk and evidence-based reasoning did not obtain. In fact, the opposite was true: students’ strongest arguments—those best supported by evidence—tended to have less conceptual substance, while conceptually richer claims and attempted explanations were more often weak or invalid scientific arguments.

We further found, in an analysis of the kinds of talk and reasoning that prevailed within the coded categories, evidence that the form in which phenomena were made available to students (numbers vs. qualitative features) was a likely contributor to the divergence between talk and reasoning we observed. Using this same detailed analysis, we also showed that ill-supported student arguments can be potentially productive conceptions, and we argued that the rigid requirement to reason from evidence inherent in scientific argumentation is a potential stumbling block for teachers and students attempting to work with such ideas in the formative assessment process.

The Compatibility of Conceptual Talk and Reasoning from Evidence

Our coded data reveal that more explicit conceptual talk did not go hand in hand with better reasoning from evidence. This result must be qualified, however, by looking at the kinds of reasoning and talk that tended to prevail in each category. Students’ less explicit conceptual talk was generally about the numerical measurements of mass and volume, while more explicit
talk—whether or not it amounted to solid reasoning from evidence—turned on qualitative characteristics and suggestive features of artifacts.

A conservative interpretation of these data would be that our observed result, the divergence of rich conceptual talk and evidence-based reasoning, was likely to have resulted from an idiosyncrasy of the formative assessment design by which qualitative and numerical data were mixed together in the same discussion. As this interpretation would have it, the provision of mass and volume as abstract numbers focused students on a conceptually empty yet evidence-supported “compare mass to volume” decision rule. Indeed it was this rule that accounted for virtually all of the evidence-supported, lower level conceptual talk—one half of the divergence story.

On the other hand, the corresponding aspect of the divergence between talk and evidence cannot be ascribed to the presence of numerical forms of data. In the twenty-four reasoning units within which students did engage in more explicit conceptual talk, appealing to numerical figures in almost none of these, they still constructed poorly-supported arguments twice as often as well-supported ones (16 vs. 8 reasoning units). This result suggests that conceptually explicit talk was difficult to combine with evidence-based reasoning, even when drawing on qualitative evidence. However, it must be pointed out that poorly supported arguments comprised a similar preponderance of conceptually implicit talk (25 vs. 16 reasoning units). Thus the fact that well-supported, explicit talk was rare (only 8 reasoning units) might not reveal an interaction of conceptual talk and reasoning from evidence but instead reflect an overall tendency of students, as a result of whatever cause, not to engage in evidence-based reasoning.

For both of the reasons outlined above, we do not conclude from our data that rich conceptual talk and reasoning from evidence should tend to be incompatible, even in
instructional contexts very similar to our own. However, our data do clearly show that these processes can be divergent; that the inclusion of numerical data was a likely factor in this divergence, and that productive student thinking was often poorly supported by evidence and therefore would have been difficult to work with in the argumentation context we described.

Bearing in mind the potentially restrictive nature of scientific argumentation, we should not lose sight of the fact that there are other forms of discussion which can incorporate rigorous scientific reasoning yet extend beyond this reasoning to embrace more diverse student thinking. The same Predict-Observe-Explain format (NRC, 2007; White & Gunstone, 1992) that is cast as scientific argumentation in the context of this study is normally a more flexible format that can readily accommodate preconceptions, anecdotes, analogies, stories, etc. that are important to students’ thinking, and which can be incorporated into student explanations, but which do not amount to valid reasoning from evidence.

It is not our purpose here to say that scientific argumentation might be better or worse than other forms of science discussion. The point we wish to make, rather, is that scientific argumentation and science discussion should be thought of as different things. From the standpoint of the teacher in the classroom, we should distinguish between a discussion centered on “what do we think is happening here?” and one that conforms to the more rigorous question of “what can we say that we now know, based on the evidence?”

*Toward a Theory of Conceptual Learning in Discussion*

What is wanted is a theory of how the process of conceptual learning can be promoted within discussion and discussion-based formative assessment. The theory we have considered here is two-fold: (1) that students construct knowledge in discussion as they articulate, reflect on, and consequently refine their emerging understandings of phenomena and (2) one way to
achieve the necessary refinement is for students to debate the meaning of evidence as scientists are known to do. We have taken some initial steps toward refining the second component of this theory. First, we have shown that conceptual talk and evidence-based reasoning may diverge, so that students’ active sense-making process is disjoined from the consideration of evidence which may constrain it. Second, we have pointed out that there is a potential for such disjunction to be exacerbated by adopting a discourse format, scientific argumentation, which necessarily restricts students claims about natural phenomena to the rigors of empirically valid reasoning, reasoning which can be at odds with imaginative or conjectural sense-making.

There are many ways to incorporate direct experience of phenomena into classroom discussion, and students’ taking on the role of scientists to reason from evidence is but one of them. If research is to provide useful guidance for curriculum designers and teachers in this area, various modes of classroom discourse appropriate for science learning need to be better distinguished from one another, and their affordances and constraints need to be the subject of empirical study. By doing so we could begin to see, for instance, when a genre like scientific argumentation might be more or less appropriate than another form. The field of science education also needs an overarching theory with which to organize various ways of framing students’ experiences of phenomena within discussion. Again, the notion that students’ experience will result in “evidence,” and that students will learn by reasoning from that evidence, is but a single framing.
References


Author Note

The data reported in this paper was collected as part of a larger study funded, in part, by a grant from the National Science Foundation (NSF/Award ESI-0095520) and, in part, by the National Center for Research on Evaluation, Standards, and Testing (CRESST/Award 0070 G CC908-A-10).