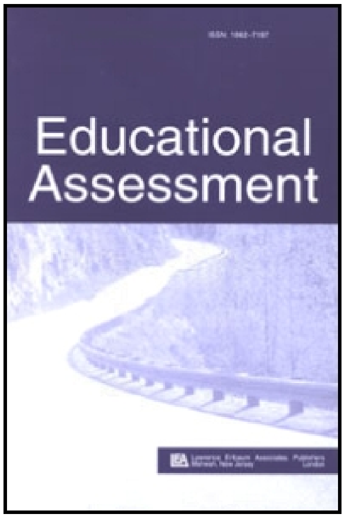

Educational Assessment
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653641

The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning

Nathaniel J. S. Brown[a]; Erin Marie Furtak[b]; Michael Timms[c]; Sam O. Nagashima[d]; Mark Wilson[e]
[a] Indiana University, Bloomington [b] University of Colorado at Boulder, [c] WestEd, [d] University of California, Los Angeles [e] University of California, Berkeley

Online publication date: 10 December 2010

To cite this Article Brown, Nathaniel J. S. , Furtak, Erin Marie , Timms, Michael , Nagashima, Sam O. and Wilson, Mark(2010) 'The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning', Educational Assessment, 15: 3, 123 — 141

To link to this Article: DOI: 10.1080/10627197.2010.530551
URL: http://dx.doi.org/10.1080/10627197.2010.530551

Educational Assessment, 15:123–141, 2010

ISSN: 1062-7197 print/1532-6977 online
DOI: 10.1080/10627197.2010.530551

Routledge
Taylor & Francis Group

The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning

Nathaniel J. S. Brown
Indiana University, Bloomington

Erin Marie Furtak
University of Colorado at Boulder

Michael Timms
WestEd

Sam O. Nagashima
University of California, Los Angeles

Mark Wilson
University of California, Berkeley

Recent science education reforms have emphasized the importance of students engaging with and reasoning from evidence to develop scientific explanations. A number of studies have created frameworks based on Toulmin's (1958/2003) argument pattern, whereas others have developed systems for assessing the quality of students' reasoning to support their scientific explanations. This article presents the centrepiece of this special issue, the Evidence-Based Reasoning Framework, which combines these two approaches to create an analytic tool intended as a foundation for assessing students' ability to reason from evidence in writing and classroom discussions. The article reviews previous frameworks developed to assess students' ability to reason scientifically and describes the elements of the Evidence-Based Reasoning Framework. It then provides an overview of the four articles in the special issue, each of which presents an application of the framework.

The creation of learning environments that develop students' abilities to reason from evidence and participate in scientific argumentation is considered a major priority in science education reform (American Association for the Advancement of Science [AAAS], 1993; National

Correspondence should be sent to Nathaniel J. S. Brown, 201 North Rose Avenue, Room 4020, Bloomington, IN 47405-1005. E-mail: njsbrown@indiana.edu

Research Council [NRC], 1996, 2001, 2007). 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). Similarly, Driver, Asoko, Leach, Mortimer, and Scott (1994) stated,

> Learning science involves young people entering into a different way of thinking about and explaining the natural world; becoming socialized to a greater or lesser extent into the practices of the scientific community with its particular purposes, ways of seeing, and ways of supporting knowledge claims. (p. 8)

This learning process must be supported by the teacher, who should seek to provide both the "appropriate experiential evidences and to make the cultural tools and conventions of the science community available to students" (Driver et al., 1994, p. 7).

To participate in arguments about scientific ideas, students must learn how to evaluate and use evidence. That is, apart from what they may already know about the substance of an assertion, students who are scientifically literate should be able to make judgments based on the evidence supporting or refuting that assertion. The use or misuse of supporting evidence, the language used, and the logic of the argument presented are important considerations in judging how seriously to take an assertion or hypothesis (AAAS, 1993, p. 298). Therefore, to be well-balanced and intelligent consumers of scientific information, students must understand that there should be ample evidence present to determine whether hypotheses are valid. This critical thinking skill is crucial for students as they develop more complete understandings of the natural world around them.

Although accumulated research from developmental and cognitive psychology indicates that students' ability to reason scientifically and to develop conceptual insights has been underestimated (Metz, 1995), students at all levels continue to have difficulties differentiating theory and evidence (NRC, 2007). Furthermore, studies have indicated that students rarely base their arguments on evidence (Bell & Linn, 2000). Analyses of classroom discourse have revealed that unsupported student conjectures are prevalent, especially in elementary school (Newton & Newton, 2000). Even in secondary school, establishing a classroom culture of scientific reasoning seems difficult to accomplish (e.g., Osborne, Erduran, & Simon, 2004). Consequently, several concerted efforts to support student reasoning in science classrooms have been undertaken with various levels of success (Carey, Evans, Honda, Jay, & Unger, 1989; Driver et al., 1994; Kawasaki, Herrenkohl, & Yeary, 2004; Tytler & Peterson, 2005).

The articles in this special issue represent the products of a German–American research consortium with the intention of creating a framework to serve as a common theoretical foundation for analyzing scientific reasoning as it manifests in a wide range of contexts. These contexts include the writing and talk of students and teachers as they engage in activities including scientific inquiry, experimentation, explanation, and prediction. This framework is intended to serve many purposes in the elementary, middle, and high school science classroom, including: (a) supporting students' and teachers' understanding of the process of scientific reasoning; (b) modeling exemplary scientific reasoning; (c) diagnosing problems and identifying pitfalls affecting student reasoning as it develops; and (d) assessing scientific reasoning in the classroom both formatively and summatively.

We have designed and developed the Evidence-Based Reasoning (EBR) Framework, described in this article and applied as described in subsequent articles in this issue, to be a new analytic tool to meet these needs for students, teachers, and researchers. In this article, we explore the role of evidence in current science education reforms and describe previous frameworks for EBR as a foundation for our own framework. Finally, we briefly introduce each article in the special issue and explain how each article takes up the EBR Framework as an assessment tool for analyzing student reasoning from evidence in writing and in classroom talk.

THEORETICAL FRAMEWORK

The scientific endeavor is by nature based on the collection and analysis of evidence; arguments based on evidence form the foundation of scientific thinking (D. Kuhn, 1993). According to Hempel (1966), "no statement ... can be significantly proposed as a scientific hypothesis or theory unless it is amenable to objective empirical test" (p. 30). By this standard, all scientific ideas must be subjected to the rigor of reality and then evaluated by evidence in the form of observations made about the natural world. Evidence also pushes the scientific enterprise forward, for observations that cannot be explained by current theories compel scientists to develop better explanatory frameworks (T. S. Kuhn, 1962).

Argumentation and EBR have been among the goals of science teaching reforms for the past 50 years. Bruner (1961) argued that learning through scientific inquiry is "in its essence a matter of rearranging or transforming evidence" (p. 22), and thus evidence should play a central role in developing students' understanding of where scientific ideas come from. Similarly, Schwab (1962) posited that students should learn about the means by which truth is derived within a discipline, or the syntactical structure of a discipline. Such learning would include the collection, analysis, and interpretation of empirical evidence, and then the development of knowledge claims supported by this evidence.

Recent efforts to make students' experiences more closely mirror the activities and thinking processes of scientists (NRC, 1996) further underscore the importance of placing scientific evidence and EBR at the core of students' experiences in science classrooms (Duschl, 2003; NRC, 1996, 2007). Students should be able to use evidence to "develop and evaluate explanations that help them address scientifically oriented questions, and formulate explanations from evidence" (NRC, 2001, p. 29). The AAAS (1990) has stated that learning science should be consistent with the nature of scientific inquiry, meaning that it should begin with questions about nature, concentrate on the collection and use of evidence, including the formulation of arguments from evidence, and be situated within the context of history.

These goals are embodied in the *National Science Education Standards* (NRC, 1996) that emphasize the importance of evidence in the science classroom and delineate five Essential Features of inquiry:

1. Learners are engaged by scientifically oriented questions.
2. Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions.
3. Learners formulate explanations from evidence to address scientifically oriented questions.

4. Learners evaluate their explanations in light of alternative explanations, particularly those reflecting scientific understanding.
5. Learners communicate and justify their proposed explanations.

A unifying characteristic of the Essential Features is their focus on students evaluating and developing scientific explanations.

The Essential Features are also reflected in the *Standards'* Fundamental Understandings in Science Inquiry, which focus on what students should know about the nature of science itself; among these targeted understandings for middle school students is the statement that "scientific explanations emphasize evidence, have logically consistent arguments, and use scientific principles, models, and theories" (NRC, 1996, p. 20).

Duschl (2003) has interpreted the Essential Features as containing three transformations in scientific inquiry: (a) data to evidence, or determining if data are anomalous or count as valid evidence; (b) evidence to patterns, or searching for patterns in and generating models for data; and (c) patterns to explanations, or developing explanations on the basis of the evidence selected. Assessment of inquiry, then, should focus on students' ability to make these transformations.

The ability to reason from evidence, along with understanding the central role evidence plays in science, is a core element in the development of scientifically literate students. Bybee (2002) argued that the capability to take part in a scientific argument is related to the area of conceptual as well as procedural scientific literacy. In its assessment framework, the Organization for Economic Co-operation and Development (2003) described the role of evidence in scientific literacy as follows:

> Scientific literacy ... gives higher priority to using scientific knowledge to "draw evidence-based conclusions." ... The ability to relate evidence or data to claims and conclusions is seen as central to what all citizens need in order to make judgments about the aspects of their lives that are influenced by science. (p. 137)

Of note in the aforementioned policy and assessment documents is that evidence should not merely be collected in science classrooms; it also should be used as a basis to support scientific arguments (Driver, Newton, & Osborne, 2000). Based on the priorities placed on evidence-based argumentation in both domestic and international contexts, it seems prudent that the manner in which students argue in classrooms, and the role evidence plays in those arguments, should be a part of the agenda for educational researchers.

Prior Frameworks for EBR and Argumentation

In the following sections, we review previous studies that have developed frameworks for exploring the role of evidence in student arguments or in student reasoning. Many of these approaches build on Toulmin's (1958/2003) *The Uses of Argument* (e.g., Driver et al., 2000; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; McNeill, Lizotte, Krajcik, & Marx, 2006; Simon, Erduran, & Osborne, 2006), and others make distinctions based on the quality of reasoning (Carey et al., 1989; Driver et al., 1994; Kawasaki et al., 2004; Tytler & Peterson, 2005). We begin by describing Toulmin's argumentation pattern and then review the previous frameworks that were influential in the development of our current framework.

Toulmin's Uses of Argument. To determine when and how evidence-based arguments are taking place, many studies have turned to Toulmin's (1958/2003) foundational framework for the analysis of arguments. Toulmin described an argument, in its basic form, as the relationship between a *claim* and the information that supports or proves the claim. Toulmin called this information *data*. Data are a sensible term colloquially but leads to confusion in the context of scientific reasoning in which the term has a specific meaning. The explanation for why the claim follows logically or causally from the data is called the *warrant*. As the warrant is a statement of a relationship, that truth of that relationship can be supported or proved by additional information called the *backing*. The relationship among these elements of an argument is represented in Figure 1.

Toulmin developed his framework to describe argument as a general means of staking and defending claims in nontechnical discourse. Reflecting this emphasis, researchers have faced difficulty in reliably identifying the elements of an argument within classroom discourse, forcing them to limit their application of Toulmin's framework (e.g., Erduran, Simon, & Osborne, 2004).

Analyses based on Toulmin's argument pattern. As previously mentioned, a number of studies have used Toulmin's argument pattern as a basis for coding systems and analytic frameworks (Sampson & Clark, 2008). These frameworks were used to analyze a variety of sources of data about student reasoning, including written assessments, classroom or small-group discussions, and clinical interviews. We summarize these studies next.

Jimenez-Aleixandre et al. (2000) analyzed conversational dynamics when students were solving authentic problems using Toulmin's argument pattern and a framework for epistemic operations. Individual arguments were divided into units of analysis, then coded as either "doing the lesson" or "doing science" (i.e., participating in scientific culture). For "doing science," the authors performed two analyses: argumentative operations and epistemic operations. For the argumentative analysis, they used a reference argument pattern consisting of data connected to a claim supported by warrants and backing. For the epistemic analysis, they drew categories from the history and philosophy of science and classroom conceptual ecology. Categories included epistemic induction, deduction, classifying, definition, causality, and plausibility, among others. Jimenez-Aleixandre et al. used their framework to diagram students' arguments according to Toulmin's pattern and to qualitatively analyze student contributions to the arguments. Findings indicated that most arguments were of a "doing school" nature, and that when students were "doing science," they developed a number of different kinds of arguments, some more sophisticated in terms of Toulmin's pattern than others. The authors also found that claims were the most common form of argument in both small-group and whole-class discussion, and backings were almost always implicit rather than explicit. Difficulties that emerged were

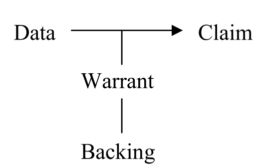

FIGURE 1 Toulmin's (1958/2003) framework for arguments (*rebuttal* omitted).

that the authors felt that some exchanges were not easily captured by Toulmin's pattern, thus creating the need to have a set of codes for epistemic operations. In addition, the authors identified the issue of determining what counts as explanation, warrant, and data, as well as previous knowledge and everyday experience.

Osborne et al. (2004) developed a framework based on Toulmin's argument pattern exploring episodes of "oppositional analysis and dialogic argument" (p. 1007) in classroom discussions. They categorized these episodes according to the manner in which claims were opposed, including elaboration, reinforcement with data and/or warrants, advancing claims, or adding qualifications. These oppositional episodes were then analyzed according to Toulmin's argument pattern. Finally, the episodes were given an additional code for their argumentative quality, ranging from Level 1, in which simple claims versus counterclaims were made, to Level 3, in which there is a series of claims or counterclaims with backing and the occasional rebuttal, to Level 5, in which there are extended arguments with more than one rebuttal. Osborne et al. reported frequencies of different argument levels through the course of a school year.

McNeill et al. (2006) also used Toulmin's argument pattern to explore how students constructed written explanations in response to fading instructional scaffolding. McNeill et al. scored students' written explanations according to a rubric that separated students' explanations into claims, evidence, and reasoning. Each of these explanation components was then divided into three levels, with Level 1 representing absent or inappropriate components, Level 2 representing accurate but incomplete components, and Level 3 representing accurate and complete components. This coding system helped the authors to create reasoning scores that could be tracked across the unit. Results indicated that students' reasoning scores were lower than their claim and evidence scores and that students' content knowledge was linked to their ability to construct scientific explanations.

In an exploration of how students construct explanations about natural selection in introductory biology courses, Sandoval (2003) identified two aspects of quality in student explanations: articulation of causal components and warrant. In this way, Sandoval distinguished students' ability to state the theory of natural selection and their ability to make sense of a particular set of data. Each student claim was then mapped onto a causal chain to explore how students were creating networks of propositions (claims). The network was then translated into a ratio of propositions in the causal chain to the total number of propositions in the network, yielding a score between 0 and 1 to indicate the relative coherence of students' explanations. Then, Sandoval evaluated the extent to which students provided data in support of each claim. Results of the analysis indicated that students did not always provide data in support of their claims when making explanations, even when their explanations were accurate. In a later study, Sandoval and Millwood (2005) evaluated students' use of evidence in written work. They focused not only on the conceptual quality but also the sufficiency of the evidence cited by students and the extent to which they used that evidence in rhetorical or persuasive arguments. Their coding system was embedded in the content of natural selection. Again, Sandoval and Millwood's findings indicate that students were able to make accurate claims about natural selection, but their ability to cite sufficient data to support their claims was dependent upon the aspect of natural selection they were arguing about.

Other frameworks for scientific reasoning and argumentation. Another set of studies in the literature explores the ways in which students develop and support scientific explanations in a variety of settings. These studies do not draw on Toulmin's argument pattern directly, but rather make distinctions on the quality of student reasoning based upon a number of factors. These studies are summarized next.

Carey et al. (1989) developed a coding system for epistemologies of science. Students participated in individual clinical interviews conducted before and after a unit about the nature of science. The interviews were then coded according to a three-level coding scheme: Level 1 responses involved no clear distinction between ideas and activities; Level 2 responses involved a clear distinction between ideas and experiments, namely, that an experiment has the purpose of testing an idea to see if it is right; and Level 3 responses had not only the distinction between Level 1 and 2 but also appreciated the relationship between the results of an experiment and the idea being tested. The interview involved questions about what science is all about, scientists' ideas, hypotheses, experiments, and results. Carey et al. used their framework to score student responses to pre- and postinterview questions and then conducted statistical comparisons between those mean scores. The coding system helped the authors to see that their intervention caused the greatest impact on the guiding ideas and questions, results and evaluation, and relationship sections of their interview. However, the authors noted that questions remained regarding making distinctions between levels, and how those levels relate to students' epistemologies. In addition, the authors argued that the role of experiments in theory building should be better connected to the process skills involved in experimentation and the relationship between students epistemologies and their conceptual understanding.

Following Driver et al.'s (1994) characterization of student reasoning in three epistemic categories, Kawasaki et al. (2004) parsed elementary students' discourse about theory and modeling into three categories for scientific inquiry: phenomenon-based reasoning, in which students make no distinction between phenomena and explanations; relation-based reasoning, or inductive relationships that are correlations between variables; and model-based reasoning, or the evaluation of theories and models on the basis of evidence. The authors used the codes to qualitatively illustrate how students developing scientific understandings transitioned between the three categories in the context of discussions about sinking and floating.

Kelly and Bazerman (2003) analyzed rhetorical moves, relationships between assertions and empirical data, and lexical cohesion in students' written arguments. Rhetorical moves were coded within sections of a scientific paper (e.g., abstract, methods, conclusion) to determine students' adherence to scientific argument in published form. Individual sentences were then taken as the unit of analysis for coding six epistemic levels along a continuum from data-pointing claims to general theoretical claims, with an additional category for students' references to personal experience or meta-level comments made by the student. These levels were subject-matter specific. A third level of analysis was performed to determine the extent to which students' arguments cohered. The authors concluded that this coding approach helped them to identify hierarchies of arguments within the papers analyzed, links across sentences to form an argument, and—of particular interest to the present analysis—the extent to which the epistemic status of students' claims varied according to the structure of the paper itself (e.g., more general claims were made in introductions and conclusions). In a separate study, Kelly and Takao (2002) critiqued these epistemic levels as missing the inferential connections made by students, leading to the possibility of mischaracterizing students' arguments as being supported by evidence when inferences are actually not being made. Furthermore, Kelly and Takao argued that they themselves may have made unwarranted inferences about students' claims and suggested that interviews with students could help researchers better understand students' uses of evidence.

Tytler and Peterson (2005) explored elementary students' reasoning in interviews with a series of four codes: the way students connected explanations and evidence, the extent to which students processed or generalized beyond the data with which they were working, how students responded to competing knowledge claims, and how students responded to anomalous data. These codes were applied to individual exchanges within an interview transcript with students. With respect to the way that students coordinated explanation and evidence, Tytler and Peterson identified three levels: Level 1 involved no systematic observations or comparisons but explanations driven by single data points, Level 2 involved making inferences about relations between variables or theoretical ideas driven by data with some conceptual interpretation, and Level 3 involved relations between variables led by theory. With respect to students' generalizations, Level 1 processing involved simply describing phenomena, Level 2 involved identifying patterns in data, and Level 3 involved making explanations. After using these levels to qualitatively illuminate the extent to which students' reason in a variety of ways, the authors argued that students' relational reasoning should be further disaggregated into three epistemological categories: phenomenon-based reasoning, relation-based reasoning, and concept-based reasoning.

In a recent review, Sampson and Clark (2008) completed a comprehensive review of analytic frameworks used to describe and assess students' scientific arguments. They distinguished among three issues critical in the study of argument in science: the structure and complexity of the argument (e.g., its components); the content of an argument (e.g., its accuracy); and the nature of the justification (e.g., how claims are backed up within an argument). They reviewed two domain-general frameworks and four domain-specific frameworks and concluded that most of the frameworks described the structure of an argument in terms of claims and justification, with claims described as assertions or explanations. They suggest that approaches emphasizing structure "clearly offer fantastic affordances in terms of providing templates for instruction and analysis that can be applied across wide ranges of contexts" (p. 466). They continued that future research should examine the connection between structural components and their relevance, sufficiency, and accuracy, as well as the epistemic nature of the justification strategies.

Rationale for a New Framework

Although we acknowledge that the previous summaries do not constitute an exhaustive review of the extensive literature on analyzing student reasoning in the context of science education, we argue that they do illustrate the variety of ways in which student reasoning and argumentation have been assessed. These studies raise a variety of extant issues regarding the assessment of student argumentation, including the difficulties in capturing the nuances of student reasoning with Toulmin's idealized framework (Jimenez-Aleixandre et al., 2000), identifying "what counts" as evidence, data, and backing (Jimenez-Aleixandre et al., 2000; Sampson & Clark, 2008), determining the extent to which researchers' inferences about the way students reason from evidence are accurate (Kelly & Takao, 2002), determining the extent to which reasoning is—or should be—content independent or content embedded (Osborne et al., 2004; Sampson & Clark, 2008; Sandoval & Reiser, 2004), and the relationship between students' conceptual understanding and ability to reason (McNeill et al., 2006).

We also identified a number of consistencies within studies attempting to identify the epistemic quality of the support students provide for their arguments. Carey et al. (1989), Driver et al. (1994), Kawasaki et al. (2004), and Tytler and Peterson (2005) all presented systems for differentiating among types of backing students provide for their claims, roughly falling into categories in which students support their claims based on individual phenomena or data points, correlations or comparisons between two variables, or generalized theories or models.

Given the ongoing challenges just identified, along with those articulated in Sampson and Clark's (2008) review, we chose not to apply Toulmin's framework directly to scientific arguments. Instead, we simplified Toulmin's framework and then adapted it to incorporate what is currently known about the process of scientific inquiry. In particular, we have synthesized the frameworks in Toulmin (1958/2003) and Duschl (2003) to create a framework of scientific reasoning as a distinct mode of thought and discourse with roots in both general argumentation and scientific inquiry. Thus, we embed certain epistemic operations within the argument pattern itself, which in turn facilitates qualitative judgments about the quality of evidence and data. The following section presents and explains the EBR Framework.

A FRAMEWORK FOR SCIENTIFIC REASONING

Our framework for scientific reasoning, called the EBR Framework, is shown in Figure 2. This framework is similar to a flowchart showing how two inputs, a premise and data, are processed through three distinct steps (analysis, interpretation, and application) to produce a claim as the output. The EBR Framework is a description of using theoretical statements,

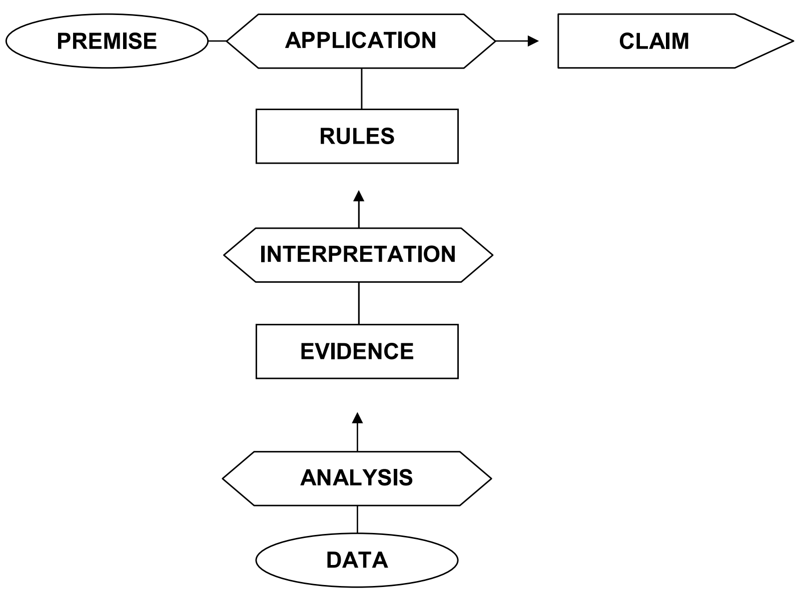

FIGURE 2 The Evidence-Based Reasoning Framework.

backed by scientific evidence, to evaluate the quality of a claim. It is not intended to model how scientific knowledge is or should be generated by students or scientists. Such knowledge-building activities may involve data, evidence, and rules, but we are not claiming that such activities are necessarily inductive in the manner suggested by the directionality of the arrows in Figure 2.

A *claim* is a statement about a specific outcome or state phrased as either a *prediction* of what something will do in the future (e.g., "This box will sink"), an *observation* of what something has done in the past (e.g., "This box sank"), or a *conclusion* about what something is in the present (e.g., "This box sinks"). It is specific to the single set of circumstances defined by the premise.

The *premise* consists of one or more statements describing the specific circumstances acting as an input that will result in the outcome described by the claim. In simple cases, as would be generally observed in a science classroom, the premise often identifies an object and a relevant feature or property (e.g., "This box is heavy"). In Toulmin (1958/2003), the same component was called *data*, a term we reserve for a different part of the Framework given its particular meaning in science.

The *rules* are the link between the premise and the claim, justifying how the latter follows from the former. Rules are statements describing a general relationship (e.g., "Something that is heavy will sink"). These relationships are general in the sense that they are expected to hold even in contexts and circumstances not previously observed. Toulmin (1958/2003) referred to this component as the *warrant*, whereas Duschl (2003), reflecting a focus on scientific inquiry, used the phrase *patterns and models*.

A rule is not necessarily correct or incorrect, scientific or intuitive. Depending on the particular instance, a rule may be an accepted scientific law such as Archimedes' Principle (i.e., "Any object immersed in a fluid is buoyed up by a force equal to the weight of the fluid displaced by the object") or an aspect of an intuitive theory (Gopnik & Wellman, 1994) such as a student's conception of buoyancy (e.g., "Heavy things sink"). The latter is both underspecified and limited in scope compared to the former. Some would call it a misconception; we prefer to identify it as a useful preconception for developing a more scientific understanding (Smith, diSessa, & Roschelle, 1994). In general, rules are intended to denote relationships between things or ideas; such relationships might consist of theories, principles, laws, propositions, correlations, or concepts of either scientific or informal aspect. Likewise, such a relationship might be described as certain, conditional, probable to a specific degree, or hypothetical. Rules in our framework are defined by their function—relating things or ideas—rather than their content or form.

Application is the process by which the rules are brought to bear in the specific circumstances described by the premise. It establishes the probability or necessity of the claim given the information described by the premise and the relationships described by the rules. In simple cases, this is often by an informal deductive logic: "This box is heavy, heavy things sink, therefore this box will sink." In more complex situations, with multiple rules governing the relationships between components, more sophisticated forms of application such as systems analysis may be necessary. The possibility of such complexity is implied by Duschl's (2003) Transformation #3, in which patterns and models are transformed into explanations.

The top region of the EBR Framework—applying the rules to the premise to produce the claim—bears a strong resemblance to Toulmin's (1958/2003) framework of argumentation.

From this point down, the resemblance will switch to Duschl's (2003) framework of scientific inquiry. This is no accident; the EBR Framework describes scientific reasoning as a two-step process in which a uniquely scientific approach to gathering and interpreting data results in rules (theories, laws, etc.) that are applied within a general framework of argumentation in which claims are justified. In the EBR Framework, the rules component is the pivot point between these steps.

Evidence consists of statements describing observed relationships. For example, the statement of evidence "the heaviest blocks sank and the lightest blocks floated" describes a relationship between the weight of the blocks and their sinking behavior. At this locus further down the Framework, additional specific context is added, making it less appropriate to provide succinct examples. In this case, there is an unspoken reference to a particular experiment conducted in a particular science classroom. The heaviest blocks are actual objects that were dropped into water and observed at a unique time and place. In the science classroom, a statement of evidence such as the one previously presented is sometimes called a "result," distinct from the observations on which it was based (see *data* next) and the principle it is intended to illustrate (see *rules* previously).

Because the evidence is grounded in a specific context, the process of *interpreting* the evidence to produce the rules is first and foremost a process of generalization. To make use of gathered evidence, it must be transformed into a statement with enough generality that it can be applied in a new situation. In the science classroom, where experiments are most often used to illustrate rather than derive a principle (Carey et al., 1989; Hart, Mulhall, & Berry, 2000), interpretation is often a crude form of overgeneralization: "These heavy blocks sank therefore buoyancy depends on mass." In more sophisticated examples of reasoning, interpretation requires weighing more than one piece of evidence including some that might serve as counterevidence. In calling this process *interpretation*, no presumption of method is made; depending on the circumstance of reasoning, it may be qualitative or quantitative, rigorous or implicit.

Although it sits in the same relative location, *evidence* is markedly different from what Toulmin (1958/2003) called the *backing*, and the process of *interpretation* has no equivalent in that framework. Although evidence and rules largely refer to the same ideas and differ primarily in their degree of generalization, backing is rarely a more specific version of the warrant. Instead, the backing introduces new ideas that correlate with or cause the warrant.

In the EBR Framework, *data* are discrete reports of past or present observations (e.g., "Block #1 sank"; "My toy boat floats in my bathtub"). Once again, these reports are deeply embedded in context; for example, Block #1 is unique and does not stand for any other block. These observations are collected and related by *analysis* to produce a statement of evidence. Here, analysis may take the form of any scientific data analysis; at the least, an analytic method involves interpolation of some form, a transformation from Block #1 and Block #2 to "these blocks." This transformation imagines a hypothetical Block #1.5 that shares the relationship expressed in the statement of evidence. Again, no presumption of method is made; depending on the circumstances of reasoning, it may be qualitative or quantitative, rigorous or implicit.

More so than the other components and processes in the EBR Framework, *data* and *analysis* retain their normal scientific meaning. This should not be surprising for, having reached the bottom of the Framework, we are firmly grounded in the principles of scientific inquiry, with no equivalent in Toulmin's (1958/2003) framework. In contrast, *interpretation* and *analysis*

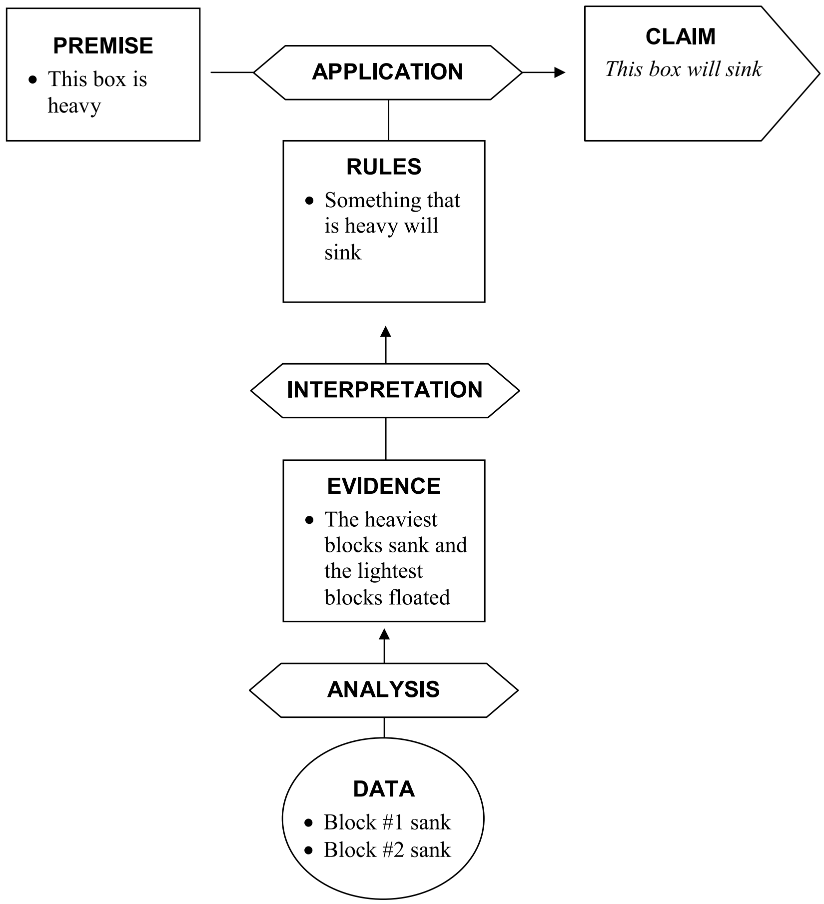

FIGURE 3 Example of the Evidence-Based Reasoning Framework as an analytic tool.

are analogous to Duschl's (2003) Transformation #2 (evidence into patterns and models) and Transformation #1 (data to evidence).

The extended but relatively simple example used throughout the previous discussion, with blocks sinking or floating, is diagrammed in Figure 3.

MAPPING TALK AND WRITING TO THE EBR FRAMEWORK

Individual Statements

The EBR Framework describes the use of evidence in scientific reasoning by students and teachers. Primarily, the framework is intended to help researchers and practitioners identify the presence and form of scientific argumentation in student work and classroom discourse. In contrast, the EBR Framework does not describe everything that teachers and students might

TABLE 1
Function of a Statement Depending on Its Relationship to Surrounding Statements

Statement	*Function in Argument*	*Surrounding Statements*	*Function in Argument*
"This block is heavy . . . "	Premise	". . . therefore it will sink"	Claim
"This block is heavy . . . "	Claim	". . . because it sank"	Premise
"This block is heavy . . . "	Half of datum	". . . and it sank"	Other half of datum

say or write in a science classroom, or even during an activity specifically designed to be an argument. Despite one's plans, such writing or discourse will inevitably contain statements not associated with the scientific argument being constructed. The EBR Framework is designed to locate and characterize the (perhaps discontinuous) elements of scientific reasoning occurring within a larger context of scientific discourse.

Individual statements are clauses consisting of a subject and predicate. We assert that a relevant statement made by a student while reasoning scientifically can be mapped to one of the components of the framework: the claim, premise, rules, evidence, or data. Because any of these components can reference a particular scientific or intuitive concept (e.g., mass), the classification of an individual statement cannot be made by semantic analysis alone. The specificity of the statement can provide some insight, as the premise, claim, and data normally reference specific objects and events (e.g., "This block sank"), whereas the rules and evidence reference more general relationships (e.g., "Heavy blocks sink"). However, specificity is ineffective in differentiating the premise, claim, and data or the rules and evidence.

Instead, it is necessary to consider the location and purpose of the statement within the context of the entire argument. In particular, one should first identify the process (i.e., application, interpretation, analysis) in which the statement functions. For example, the statement "This block is heavy" could be a premise, claim, or datum based on its specificity. The surrounding statements can illuminate which process of reasoning is being employed, as illustrated in Table 1.

Individual Arguments

We believe that sufficiently advanced students, with scaffolding, can produce statements representing all five components (premise, claim, rules, evidence, and data) for any given scientific argument. That is to say, we believe that all parts of the framework are always relevant and potentially observable. However, in the vast majority of arguments, students will make statements that represent only parts of the EBR Framework; in these cases, the existence of the other components must be inferred.

Based on the data presented in the subsequent articles in this special issue, we propose a hierarchy of incomplete arguments, illustrated in Figure 4, that represents more and more sophisticated forms of scientific reasoning. Because scientific arguments must be justified by evidence grounded in data, such a hierarchy can be useful for evaluating the strength of an argument. This hierarchy may also represent a path that students pass through as they learn to reason with evidence. This path would suggest a sequence of instruction that is perhaps more natural than the typical bottom-up approach taught in science classrooms, in which one first

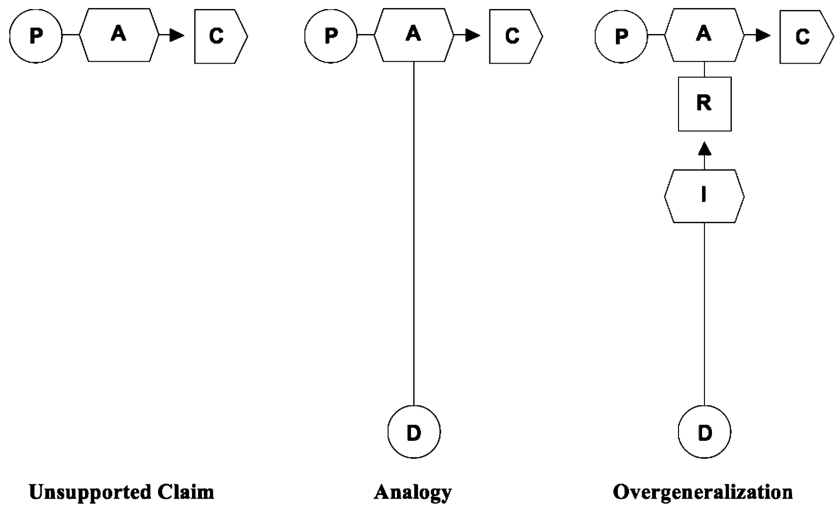

FIGURE 4 Proposed hierarchy of increasing sophistication of incomplete arguments. *Note.* P = premise; A = application; C = claim; D = data; R = rules; I = interpretation.

collects data, then analyzes the data to get a result, and finally interprets that piece of evidence to infer a rule.

There are likely to be many reasons why students may not be able to produce a full-fledged argument that employs all parts of the framework. First, they may not have developed the ability to conceptualize the different parts of the framework. D. Kuhn (1993) demonstrated that children often cannot differentiate results, theory, and evidence. In the EBR Framework, we would say these children confuse the inputs and outputs of the framework, confounding the premise, claim, and data. Second, some students may be capable of conceptualizing the differences between the parts of the framework but do not do so because they have not learned these differences (Carey et al., 1989; Driver et al., 1994; Kawasaki et al., 2004; Tytler & Peterson, 2005). That is to say, students will likely need to receive instruction on the differences between the parts of the framework before they can produce all five statements for a given scientific argument. Third, students who have learned the framework may still not produce all five statements for rhetorical reasons (Kelly & Bazerman, 2003), unless they are sufficiently scaffolded or prompted. As illustrated in the second article presented (Brown, Nagashima, Fu, Timms, & Wilson, this issue) when asked to justify a claim, some students may naturally cite the premise, other students may cite the rules, and still others may cite examples of data or evidence. Even professional scientists rarely, if ever, appeal all the way back to individual pieces of data collected in historical experiments when asked to support a claim, despite surely being able to look up and reference such data if needed.

Currently, we do not know the extent to which it is possible, in collected assessment data, to distinguish between these three interpretations of an incomplete argument. We recommend that sufficient scaffolding be used in the assessment or teacher prompts so that the effect of the third interpretation can be minimized. Even so, we have found it difficult to evaluate some responses when scaffolding has failed to produce a complete argument.

Counterevidence and Debate

The EBR Framework also provides a way of representing scientific argumentation between two or more positions. When two people are arguing, or when one person is weighing counterevidence, separate diagrams can represent each argument.

Figure 5 shows a hypothetical example of how a scientific debate between two students might be mapped to the EBR Framework. In this case, both sides are arguing from the same premise but different rules stemming from different evidence and data. This structure is a characteristic of this debate, and other debates may have a different structure. For example, the premises may be different while the data are the same. Note that Student A's data is counterevidence against Student B's data, and vice versa.

As for a single argument, the quality of these individual arguments can be judged using the hierarchy in Figure 4. When considering completeness in this manner, Student A's analogy is weaker than Student B's overgeneralized argument. Likewise, one could consider the amount of data and evidence supporting each argument. Student B's argument is grounded in more data, and therefore stronger in this sense as well.

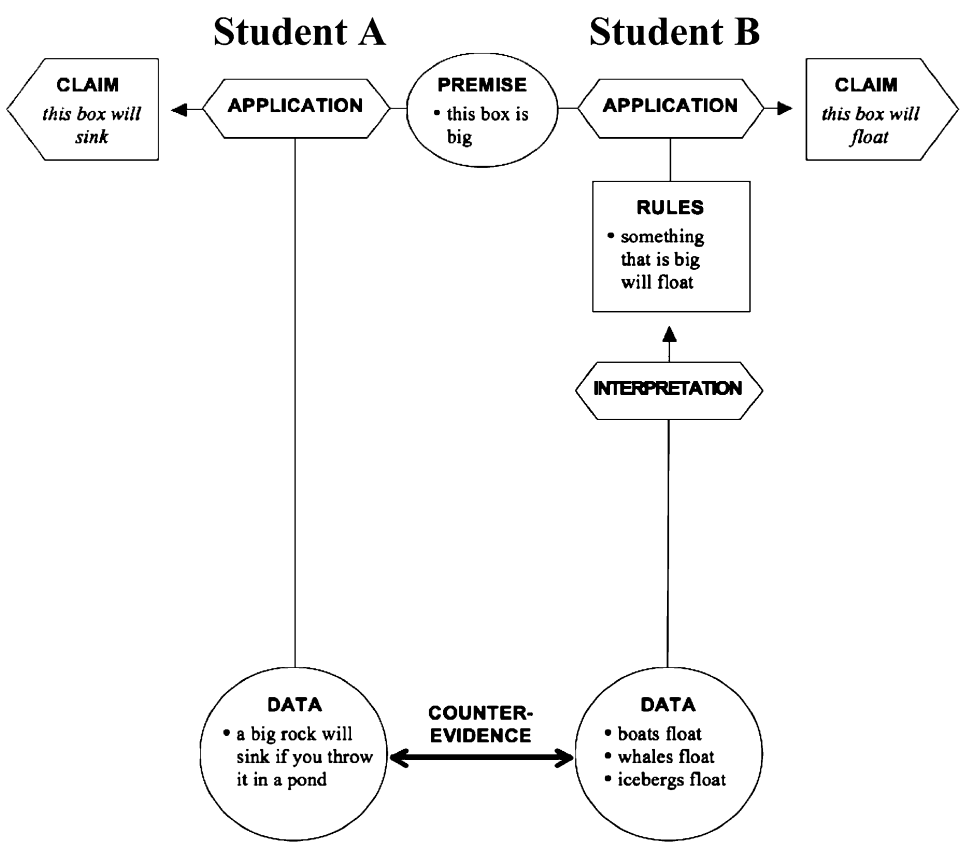

FIGURE 5 A two-sided scientific argument employing counterevidence.

APPLICATIONS OF THE EBR FRAMEWORK

The EBR Framework does not serve directly as an assessment instrument. Rather, it provides for assessment developers and practitioners a theoretical lens to help identify the elements and structure of a scientific argument. Identifying when and in what form scientific reasoning occurs is the first step in assessing its quality. By describing the individual components and processes involved in scientific reasoning, the EBR Framework allows different aspects to be selected as a focus for assessment and subsequent interpretation. Consequently, the EBR Framework is intended to establish a common foundation for the development of a wide range of assessment tools and contexts. Such individual assessments may focus on written work or classroom talk, be designed for formative or summative purposes, and encompass many formats; for some examples, see the other articles in this special issue. In general, the EBR Framework is designed to be agnostic to the type and purpose of specific assessments that are created using the Framework as a foundation.

When using the EBR Framework, we do not recommend independently assessing the accuracy or quality of the claim apart from the quality of the components and processes making up the supporting argument. Although it is tempting in any scoring rubric to include the simple judgment of whether the claim is correct, we believe doing so runs the risk of short-circuiting the evaluation of the argument in the minds of both students and teachers. When using the EBR Framework, one has already committed to a deeper and more nuanced assessment of the scientific reasoning and argumentation that gives rise to the claim.

For example, as statements of the principle or concept a student is using, the rules are an attractive locus for assessment. Although the assessment of the rules could be reduced to a correct/incorrect dichotomy, we prefer to consider the relative conceptual sophistication of the rules, ranging from preconceptions to scientific conceptions. As an example, the rule used previously ("Something that is heavy will sink") is a statement that is more sophisticated than some (e.g., "Something with holes will sink") yet less sophisticated than others (e.g., "Something that is dense will sink") even though all three are scientifically inaccurate or incomplete. Each of the articles in this special issue will deal with the conceptual sophistication of the rules in some regard.

There are many other characteristics of the rules on which one could choose to focus, including the breadth of their applicability, how reliably they are used in similar scenarios, and even their grammar and spelling. The power of the EBR Framework is that it supports multiple goals and allows assessment developers and practitioners to focus on what they deem important.

Overview of the Special Issue

We hope that this analytic framework for the use of evidence in scientific reasoning will form a robust platform for developing research and assessment instruments that can detect and measure students' and their teachers' use of evidence in science teaching. The utility of the framework was tested by a collaborative team of researchers from research institutions and universities in the United States and in Germany in a series of three studies, reported in the following articles.

In the second article (Brown, Nagashima, Fu, Timms, and Wilson, this issue), the EBR Framework was used to develop highly scaffolded written test items designed to probe students' use of evidence. The items were administered to middle school and high school students, and several variables were used to evaluate the quality of their arguments. In addition to conceptual sophistication, the specificity of the rules was also a focus. As an example, the statement "Something that is dense will sink" is more specific than the statement "Sinking depends on density" and less specific than the statement "Something that is denser than 1.0 g/mL will sink." All three statements reflect the common concept of density but become progressively more specific and quantitative. The processes involved in scientific reasoning were likewise a focus of assessment. In particular, the validity of each process (analysis, interpretation, application) was characterized, where validity is defined as whether the result follows logically from the assumption, regardless of the accuracy of the assumption. For example, given the rules and the premise, does the claim follow as a logical necessity? If so, the application is said to be valid, even if the rules, and hence the claim, are inaccurate. It was found that, using appropriate scaffolding in the item stems, item sets could be designed that collectively elicit all of the components and processes of the EBR Framework. Thus, it was possible to draw out and assess the nature and quality of students' use of evidence using a sequence of written assessment items.

The third article (Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, this issue) presents an adaptation of the EBR Framework to create the Argumentation in Science Classroom Discourse instrument. This adaptation was created to measure the quality of student reasoning in whole-class discussions and to capture teachers' and students' co-constructed reasoning about scientific phenomena. The article explains the ways in which the EBR Framework was adapted as a video analysis tool, explains the sets of codes that compose the tool, and then presents sample coded classroom discourse to illustrate how the tool functions. The fourth and fifth articles in this issue then apply the Argumentation in Science Classroom Discourse tool to two data sets consisting of students reasoning in classroom discussion. Both data sets are based on instructional variations designed to promote students' understanding of sinking and floating, one involving different degrees of scaffolding, the other involving prior instruction on nature-of-science constructs.

In the fourth article, Hardy, Kloetzer, Moeller, and Sodian (this issue) apply the Argumentation in Science Classroom Discourse tool to analyze two data sets from six third- and fourth-grade science classrooms. Findings of the study indicate that the majority of discourse involved unsupported claims about the scientific phenomena, although there was some evidence for effects of a preceding curriculum that had focused on nature-of-science constructs on the quality of students' EBR. Hardy et al. also gave conceptual codes to student dialogue and found correlations between students' conceptual understanding and reasoning level within reasoning units. Finally, they found that teacher prompts for providing support for conclusions and inferences were associated with higher reasoning levels, emphasizing the role of teachers in promoting a culture of productive use of evidence in classroom discourse.

In the fifth article, Shemwell and Furtak (this issue) use the Argumentation in Science Classroom Discourse tool to examine the suitability of scientific argumentation, a discourse format that gives priority to reasoning from evidence, as a form of classroom discussion for conceptual learning. Analyzing scientific argumentation sessions from six different middle school classrooms, they found that segments of discussion in which students back up their claims with evidence often fail to incorporate descriptive elaboration of concepts or causal mechanisms. Conversely, more elaborated description occurs most often as claims or with claims that are not supported by evidence. In a further analysis of coded discussion segments, the authors show how giving priority to reasoning from evidence incurs constraints that tend to hinder descriptive elaboration.

We plan to develop other applications of the EBR Framework in other science content areas, and we hope that other researchers, test developers, and curriculum developers will also apply the Framework in the creation of research instruments, assessments, and instructional materials that test the Framework's utility further. Over time we hope that this will lead to improvements in teaching students the key skills of using evidence in scientific reasoning and thereby promote growth in students' understanding of science concepts.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under Grant 0439062. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

American Association for the Advancement of Science. (1990). *Science for all Americans*. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

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

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010/this issue). A Framework for Analyzing Scientific Reasoning in Assessments. *Educational Assessment, 15*, 142–175.

Bruner, J. (1961). The act of discovery. *Harvard Educational Review, 31*, 21–32.

Bybee, R. W. (2002). Scientific literacy: Mythos oder Realität? In W. Gräber (Ed.), *Scientific literacy. Der Beitrag der Naturwissenschaften zur allgemeinen Bilduing* (pp. 21–43). Opladen, Germany: Leske und Budrich.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). "An experiment is when you try it and see if it works": A study of Grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*, 514–529.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*, 5–12.

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

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*, 37–73.

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

Gopnik, A. & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). Cambridge, England: Cambridge University Press.

Hart, C., Mulhall, P., & Berry, A. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching*, *37*, 655–675.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice-Hall.

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). “Doing the lesson” “Doing science”: Argument in high school genetics. *Science Education, 84*, 757–792.

Kawasaki, K., Herrenkohl, L. R., & Yeary, S. A. (2004). Theory building and modeling in a sinking and floating unit: A case study of third and fourth grade students’ developing epistemologies of science. *International Journal of Science Education, 26*, 1299–1324.

Kelly, G. J., & Bazerman, C. (2003). How students argue scientific claims: A rhetorical-semantic analysis. *Applied Linguistics, 24*, 28–55.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students’ use of evidence in writing. *Science Education*, *86*, 314–342.

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

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students’ construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*, 153–191.

Metz, K. (1995). Reassessment of developmental constraints on children’s science instruction. *Review of Educational Research*, *65*, 93–127.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

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

National Research Council. (2007). *Taking science to school: Learning and teaching science in Grades K–8.* Washington, DC: National Academies Press.

Newton, D. P., & Newton, L. D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? *British Educational Research Journal, 26*, 599–613.

Organization for Economic Co-operation and Development. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science, and problem solving knowledge and skills*. Paris, France: Organization for Economic Co-operation and Development.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*, 994–1020.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*, 447–472.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students’ scientific explanations. *The Journal of the Learning Sciences, 12*, 5–51.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students’ use of evidence in written scientific explanations. *Cognition and Instruction, 23*, 23–55.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345–372.

Schwab, J. J. (1962). The concept of the structure of a discipline. *The Educational Record, 43*, 197–205.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education, 28*, 235–260.

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

Toulmin, S. E. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press. (Original work published 1958)

Tytler, R., & Peterson, S. (2005). A longitudinal study of children’s developing knowledge and reasoning in science. *Research in Science Education, 35*, 63–98.