Assessment of the Ways Students Generate Arguments in Science Education: Current Perspectives and Recommendations for Future Directions

VICTOR SAMPSON
Department of Middle and Secondary Education, College of Education, Florida State University, Tallahassee, FL 32306-4490, USA

DOUGLAS B. CLARK
College of Education, Arizona State University, Tempe, AZ 85287-0911, USA

Received 10 February 2007; revised 15 November 2007, 16 January 2008; accepted 23 January 2008

DOI 10.1002/sce.20276
Published online 13 March 2008 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: Theoretical and empirical research on argument and argumentation in science education has intensified over the last two decades. The term "argument" in this review refers to the artifacts that a student or a group of students create when asked to articulate and justify claims or explanations whereas the term "argumentation" refers to the process of constructing these artifacts. The intent of this review is to provide an overview of several analytic frameworks that science educators use to assess and characterize the nature of or quality of scientific arguments in terms of three focal issues: structure, justification, and content. To highlight the foci, affordances, and constraints of these different analytic methods, the review of each framework includes an analysis of a sample argument. The review concludes with a synthesis of the three focal issues and outlines several recommendations for future work. Ultimately, this examination and synthesis of these frameworks in terms of how each conceptualizes argument structure, justification, and content is intended

An earlier version of this article, entitled "Assessment of Argument in Science Education: A Critical Review of the Literature," appeared in S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.). (2006). *Proceedings of the Seventh International Conference of the Learning Sciences—Making a Difference* (pp. 655–661). Mahwah, NJ: Erlbaum.
Correspondence to: Victor Sampson; e-mail: vsampson@fsu.edu
Contract grant sponsor: National Science Foundation.
Contract grant number: 0334199.
Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

to provide a theoretical foundation for future research on argument in science education.
 Sci Ed **92:**447–472, 2008

INTRODUCTION

Research on argumentation in science education has expanded and intensified considerably over the past two decades. Our review of this literature indicates that substantial variation exists in the analytic frameworks that have been developed to study the scientific arguments that students construct. This variation tends to reflect the goals of the research programs and the underlying theoretical perspectives about argument in science and learning that frame the research designs. Regardless of these diverse perspectives, however, it seems that these frameworks share several focal issues. In this review, we examine several analytic frameworks that science education researchers have employed to assess the nature and quality of students' arguments, highlight the constraints and affordances of each framework, and make recommendations for future investigations. It is our hope that this review will provide a theoretical foundation for planning, executing, and communicating the findings of future research examining argument in science education.

BACKGROUND ON ARGUMENT AND ARGUMENTATION IN SCIENCE EDUCATION

Research in science education has examined the artifacts that students create to articulate and justify claims, explanations, or viewpoints (e.g., Bell, 2004; Lawson, 2002; Sandoval & Millwood, 2005; Zohar & Nemet, 2002) and the processes through which groups of students engage with one another as they propose, critique, and evaluate ideas (e.g., Abell, Anderson, & Chezem, 2000; Clark & Sampson, 2006b; Kuhn & Reiser, 2006; Kuhn & Udell, 2003; Osborne, Erduran, & Simon, 2004; Veerman, 2003). Although significant overlaps exist, this article focuses primarily on the former. Throughout this review, we will use the term "argument" to describe the artifacts students create to articulate and justify claims or explanations and the term "argumentation" to describe the complex process of generating these artifacts (e.g., Osborne et al., 2004, p. 998). It is important to note that these distinctions are not meant to be absolute. These distinctions do, however, represent traditional distinctions of emphasis in science education research and allow us to focus this review accordingly. (See Clark, Sampson, Weinberger, & Erkens, 2007; Erduran, 2008, for reviews of research devoted to more dialogical interactions where individuals propose, support, and critique ideas through discussion or debate.)

THE ROLE OF ARGUMENT AND ARGUMENTATION IN SCIENCE AND SCIENCE EDUCATION

Scientific inquiry is often described as a knowledge-building process in which explanations are developed to make sense of data and then presented to a community of peers for critique, debate, and revision (Driver, Newton, & Osborne, 2000; Duschl, 2000; Passmore & Stewart, 2002; Sandoval & Reiser, 2004; Stewart, Cartier, & Passmore, 2005; Vellom & Anderson, 1999). Thus, the ability to generate a persuasive and convincing argument that coordinates evidence and theory to support or refute an explanation is an important component of the inquiry process (Driver et al., 2000; Duschl & Osborne, 2002; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kuhn, D., 1993; Kuhn, T. S., 1970; Latour, 1987; Siegel, 1989). However, in order for arguments to be considered persuasive and convincing, they must be consistent with the epistemological criteria used by the larger scientific community for "what counts" as valid and warranted scientific knowledge. Examples of important epistemological criteria in science include (a) the need to provide evidentiary backing or rationales for knowledge claims and proposed tests of claims (Hogan & Maglienti, 2001), (b) the need for coherence between theoretical frameworks and observations of phenomena (Passmore & Stewart, 2002), (c) the importance of establishing the credibility of evidence (Driver et al., 2000), (d) the value of parsimony (Sandoval & Reiser, 2004), and (e) the importance of basing arguments on reasoning that is logically valid (Zeidler, 1997).

Current research indicates that learning how to engage in productive scientific argumentation to propose and justify an explanation through argument is difficult for students. Students need to learn more about the types of claims that scientists make, how scientists advance them, what kinds of evidence are needed to warrant one idea over another, and how that evidence can be gathered and interpreted in terms of community standards (Kelly & Chen, 1999; Osborne, 2002; Sandoval & Reiser, 2004). Duschl (2008) describes these various types of knowledge as the conceptual, cognitive, epistemic, and social aspects of generating and evaluating arguments. Given this perspective, a number of science educators have argued that students need more opportunities to learn how the scientific community uses arguments to construct knowledge and the criteria for what counts as a good argument in science as part of a quality education in science (e.g., Bell & Linn, 2000; Driver et al., 2000; Duschl & Osborne, 2002; Kuhn & Reiser, 2005; McNeill, Lizotte, Krajcik, & Marx, 2006; Newton, Driver, & Osborne, 1999; Sandoval, 2003). Thus, empirical research that examines how students generate arguments and how they learn about this process has become an area of major concern for science education research. In the sections that follow we examine some of the various ways researchers describe and evaluate student-generated arguments in this context.

ASSESSING THE NATURE OR QUALITY OF ARGUMENTS IN THE CONTEXT OF SCIENCE EDUCATION

How well can students, who are not members of the scientific community, generate an argument that justifies or refutes a particular interpretation of a natural phenomenon? How do these processes and assumptions differ from those ideally employed by scientists? Are students able to assimilate the practices of argumentation valued by the scientific community as a result of classroom instruction? To answer these types of questions, researchers have developed a number of analytic frameworks for examining the nature and the quality of student-generated arguments. This selective review is designed to provide an overview of these frameworks.

We have organized this review around three issues that seem to be of critical importance to those who study the ways students generate argument in the context of science: (1) the structure or complexity of the argument (i.e., the components of an argument), (2) the content of an argument (i.e., the accuracy or adequacy of the various components in the argument when evaluated from a scientific perspective), and (3) the nature of the justification (i.e., how ideas or claims are supported or validated within an argument). These themes thus provide lenses for analyzing the theoretical perspectives underlying the frameworks, the pedagogical or research goals of the respective researchers, and the relative affordances (or constraints) of each approach for studying the arguments that students generate in the context of science education.

We begin our analysis with two relatively *domain-general* frameworks (i.e., frameworks that can be used analyze argument quality inside or outside of the field of science) and then consider four relatively *domain-specific* frameworks (i.e., frameworks that focus on aspects or criteria of argument specific to science or subfields and specific contexts within

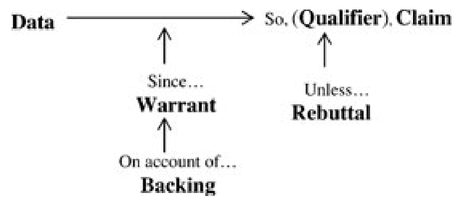

Figure 1. Toulmin's argument pattern.

science). To help highlight the various ways these frameworks define structure, content, and justification as well as the differing foci of the frameworks in terms of relative weights placed on structure, content, and justification, we analyze a sample argument using each approach. This argument was generated by a middle school student who was asked to articulate and justify an explanation for why some objects, such as a metal and a wooden spoon, feel different even though they have sitting in the same room for several hours[1]:

> I think all objects in the same surroundings become the same temperature even if an object produces its own heat energy. This is true because on the lab that we did all the temperatures were in their 20s which proves that the room temperature changes the objects to the same as the room. Therefore, even though objects may feel different, they are actually within a few degrees of each other.

Domain-General Analytic Frameworks

We first consider two frameworks that are relatively domain general in terms of their applicability and the criteria they use to analyze student-generated arguments. These frameworks afford application across a broad range of contexts and purposes but potentially forgo opportunities to consider important aspects of argumentation particular to specific disciplines or aspects of scientific endeavors. The frameworks we consider here include Toulmin's argument pattern (1958) and an approach developed by Schwarz, Neuman, Gil, and Ilya (2003).

Toulmin's Focus on the Pattern of an Argument. Toulmin's perspective on argumentation has substantially influenced science education research. In *The Uses of Argument* (1958), Toulmin distinguished between the idealized notions of logical-formal arguments as used in mathematics and the use of arguments in linguistic contexts. Toulmin's perspective on argument has enabled researchers to use this framework to examine argument quality in a variety of domains including language arts (e.g., Reznitskaya et al., 2001), economics (e.g., Cho & Jonassen, 2002), mathematics (e.g., Forman, Larreamendy-Joerns, Stein, & Brown, 1998; Krummheuer, 1995), and science (e.g., Bell & Linn, 2000; Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre et al., 2000; Kelly & Chen, 1999; Osborne et al., 2004).

Mechanics of the Framework. Toulmin's argument framework suggests that the statements that make up an argument have different functions that can be classified into one of six categories: claims, data, warrants, backings, qualifiers, and rebuttals (Figure 1). According to this framework, claims are assertions and data are the foundations for those

[1] This sample argument is taken from our own research (Clark & Sampson, 2006a), which examines how middle school students construct and evaluate arguments within the domain of thermodynamics.

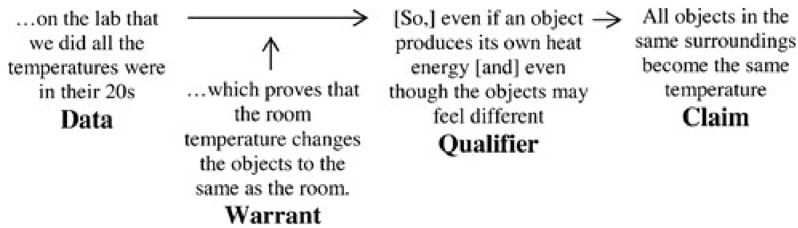

Figure 2. Analysis of the sample argument using Toulmin's argument pattern.

claims. Warrants are comments that are used justify why data are relevant to the claim. The warrant's strength is indicated by the inclusion of a modal qualifier. The backings of an argument are the comments that are used to establish the general conditions that strengthen the acceptability of the warrants so that the connection between the data and the claims will not be questioned. Finally, a rebuttal indicates the "circumstances in which the general authority of the warrant would have to be set aside" (p. 101/p. 94 in updated 2003 edition). Toulmin describes the process of constructing a scientific argument primarily as a process of using data, warrants, and backings to convince others of the validity of a specific claim. From this perspective, the strength of an argument is based on the presence or absence of specific combinations of these structural components.

Application to the Sample Argument. An analysis of the sample argument using Toulmin's argument pattern (Figure 2) indicates that this student-included data ("on the lab that we did all the temperatures were in their 20s"), a warrant ("which proves that the room temperature changes the objects to the same as the room"), and two qualifiers ("[so,] even if an object produces its own heat energy [and] even though the objects may feel different") to support their claim. This argument might, therefore, be considered strong from the perspective of this framework because it includes all of the components of a quality argument except a backing, which would be used to support the validity or acceptability of a warrant.

Synthesis in Terms of Structure, Justification, and Content. The main focus thus far of research involving *Toulmin's argument pattern* in science education contexts has focused on structural issues. Researchers who used *Toulmin's argument pattern* in their work, for example, have provided a great deal of insight into the ways students structure an argument and the nature of the justification they use to support their ideas. For example, Bell and Linn (2000) used this framework to analyze the arguments produced by students to explain the nature of light. They found that students tend to rely on data to support their claims but frequently do not include warrants or backings. Similarly, Jimenez-Aleixandre et al. (2000) found that high school students, when constructing arguments about genetics, focus on making detailed claims but do not support them with data or warrants. The findings from this type of research, as noted earlier, have had a substantial influence on the ways science educators design technology-enhanced learning environments (e.g., Clark & Sampson, 2006a; Erkens, Kanselaar, Prangsma, & Jaspers, 2003; McNeill et al., 2006; Weinberger, 2003; Zembal-Saul, Munford, Crawford, Friedrichsen, & Land, 2003) and classroom activities (e.g., Kuhn & Reiser, 2005; Osborne et al., 2004; Simon, Erduran, & Osborne, 2006) to help students learn how to engage in argumentation (i.e., argument generation) in a more productive manner.

One complication encountered by researchers in applying Toulmin's framework, however, involves reliably distinguishing between claims, data, warrants, and backings because the comments made by students can often be classified into multiple categories (e.g., Eichinger, Anderson, Palincsar, & David, 1991; Forman et al., 1998; Jimenez-Aleixandre et al., 2000; Kelly, Druker, & Chen, 1998). For example, in Figure 2, the comment "even if the object produces its own heat energy" could serve as part of the claim, or as a qualifier (indicating the strength of the warrant) or even as a rebuttal (indicating a circumstance in which the general authority of the warrant would have been set aside). Furthermore, when arguments are longer, as is the case when students are writing a journal article or a position paper, statements may serve as a new claim (thus requiring support) or as a warrant for a preexisting claim (Kelly & Takao, 2002). As a result, a researcher's personal perspectives about what should count as a warrant, claim, or data will often influence how he or she codes a comment using this analytic framework. This type of bias typically has an adverse effect on interrater reliability and has caused some researchers to question the usefulness of this framework for studying arguments generated by students in the context of science (e.g., Duschl, 2008; Duschl, Ellenbogan, & Erduran, 1999).

In terms of issues of justification and content, although Toulmin's framework stresses the importance of field dependence (pp. 33–36 of updated edition), the framework itself provides little specific information about these *field-dependent* features. According to Toulmin's framework, claims, warrants, data, and backings are *field-invariant* features of an argument that can be used to study the structure of an argument regardless of context. Toulmin explains that what count as appropriate claims, warrants, backings, or data are *field-dependent* features of an argument (Toulmin, 1958). Unfortunately, because the majority of the research using Toulmin's argument framework has focused on the field-invariant features of an argument, we know very little about how well arguments constructed by students adhere to the criteria shared by the scientific community for judging quality. For example, do students incorporate evidence that is valid and reliable as data in their argument? Do students attempt to coordinate their claim with all available data or just the data that support their particular viewpoint? Answers to these types of questions can provide valuable insights into students' understanding of what counts as a quality argument in science.

In addition, research relying on standard Toulmin frameworks has generally provided less insight in terms of other issues of justification and content. For example, although the sample student argument would be considered relatively strong structurally according to most Toulmin-based frameworks, the content is inaccurate from a scientific perspective. The comment, "even if an object produces its own heat energy," indicates that this student believes that all objects reach thermal equilibrium, when in fact an object that produces its own heat energy would be a circumstance where the general authority of the warrant would have to be set aside, thus serving as a rebuttal to the claim "all objects in the same surroundings become the same temperature." Similarly, a standard application of Toulmin's framework does not include an assessment of the logical structure and coherence of the justification beyond the presence or absence of data, warrants, and backings. Hence, all that matters is their presence or absence regardless of accuracy or relevance. As a result, those interested in examining the content of an argument must supplement this framework with other measures because it does not take into account the accuracy of the components from a scientific perspective or even if the argument, as a whole, makes sense.

Schwarz, Neuman, Gil, and Ilya's Focus on the Structure and the Acceptability of Reasons in an Argument. Another example of a relatively domain-general approach that has been used to examine student-generated arguments in the context of science education is a framework developed by Schwarz and colleagues (2003). This framework was designed for contexts where students "produce text arguments in structured interviews

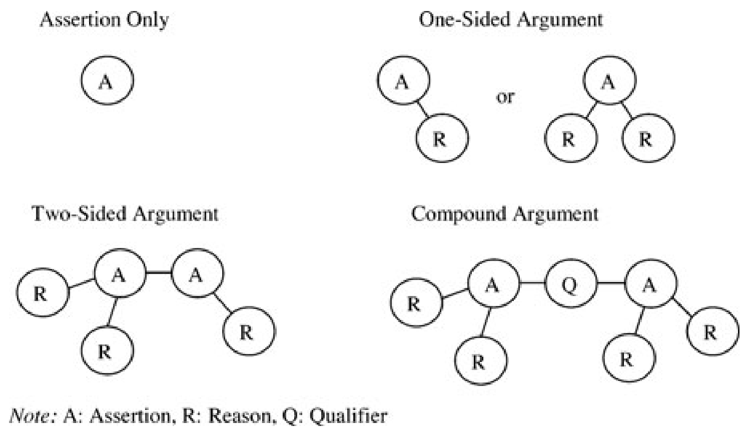

Figure 3. Possible structures of an argument in the framework of Schwarz et al. (2003).

or in essays where they were invited to express their standpoint" and "participants know they are expected to be explicit" (p. 229). This framework focuses on structural complexity and the nature of the justification to evaluate argument quality rather than content. As a result, this framework can be used in a variety of different contexts with little to no modification because, like Toulmin's, it is based on the assumption that significant aspects of the strategies used by individuals to generate a quality argument are not context dependant.

Mechanics of the Framework. Schwarz et al. define an argument simply as a conclusion with at least one reason. They explain, however, that arguments can be elaborated with qualifiers, as per Toulmin, as well as with multiple reasons, counterarguments, and metastatements. The soundness of an argument, from their perspective, involves the acceptability of the provided reasons and the relevance of those reasons. In practice, these aspects of an argument are quantified by identifying the (a) argument type, (b) soundness of argument, (c) overall number of reasons, (d) number of reasons supporting counterarguments, and (e) types of reasons included.

Schwarz and colleagues' hierarchy of argument structure is fairly simple, ranging from a *simple assertion* to a *compound argument* (see Figure 3). *Simple assertions* consist of a conclusion that is not supported by any type of justification. *One-sided* arguments include only a conclusion and one or more reasons. *Two-sided* arguments include reasons that both support and challenge the conclusion but "do not show clearly whether the student or group undertook an analysis of the pros and cons necessary to solve the issue," (p. 229). *Compound* arguments, on the other hand, make this type of analysis explicit by including such phrases as "it depends. . . , if. . . , but only if. . . ," (p. 229). Schwarz et al. also examine other factors related to structure such as soundness and overall number of reasons. Soundness of an argument is judged by examining the acceptability of the argument (based on the logical structure and its degree of realism) and the relevance of the reasons used to support a conclusion, both of which are evaluated on a scale of 0–2.

The classification of the quality of reasons used by Schwarz et al. is adapted from the work of Means and Voss (1996) and is considerably more nuanced. *Abstract* reasons are logical in form where a definition is applied to make a decision (e.g., "experiments help to develop medications," p. 232). *Consequential* reasons refer to a direct consequence of an action as a reason, for example, "it helps save lives" (p. 232). *Make sense* reasons include reasons that consist of generally accepted beliefs or truisms (e.g., "animals have feeling like humans," p. 232), appeals to authority (e.g., "because I heard a scientist on TV who

TABLE 1
The Sample Argument Coded Using an Analytic Framework Developed by Schwarz, Neuman, Gil, and Ilya

Component of the Argument	Code
I think. . . all objects in the same surroundings become the same temperature even if an object produces its own heat energy.	Assertion
This is true because on the lab that we did all the temperatures were in their 20s, which proves that the room temperature changes the objects to the same as the room.	Abstract reason
Therefore, even though objects may feel different, they are actually within a few degrees of each other.	Metastatement

said that it helps fight disease," p. 232), or reasons that are based on personal experiences (e.g., "because I once dissected a frog and its appalling," p. 232). Finally, *vague* reasons consist of imprecise statements (e.g., "because the animal can be out of control and hurt the scientists and the scientists will be sick and there will be an epidemic," p. 232).

Application to the Sample Argument. Table 1 shows how Schwarz and colleagues' analytic framework would be applied to the sample argument. The sample represents a one-sided argument in terms of structure because it includes an assertion, a reason, and a metastatement. The argument is sound from the perspective of this framework because the reason is acceptable and relevant to most of the claim. One possible challenge to this interpretation of "acceptable" and "relevant" is that the reason does not support or acknowledge the "even if the objects produce their own heat energy" clause of the conclusion. Overall, this sample argument would not be considered as strong as it would when analyzed using *Toulmin's argument pattern* because it is one sided and the assertion is only supported by one reason (although this reason is acceptable, relevant, and strong).

Synthesis in Terms of Structure, Justification, and Content. The Schwarz et al. framework, as noted earlier, focuses primarily on issues of structure and justification and less on the issue of content. This focus has provided additional insight into the argumentation practices of students that can be overlooked with an analysis that relies on Toulmin's argument pattern. For example, the work of Schwarz and colleagues (2003) suggests that students tend to produce one-sided arguments that consist of an assertion supported by a single make-sense or vague reason. However, after participating in an intervention that consisted of cycles of reading, discussion, and argument generation with a classmate, the students were able to generate arguments that were less one sided, more compound, and included reasons that were more acceptable and relevant when working on their own. This work suggests that a lack of familiarity with an issue (in this case animal research) can prevent students from producing an argument with a complex structure (because they do not understand or they are unaware of the opposing viewpoint) and an acceptable justification (because they do not know enough about the topic to provide abstract or consequential reasons).

However, it is important to note that this approach, like Toulmin's, does not use domain-specific criteria to assess the quality of the justification. In this framework, the highest quality reasons are *abstract* (e.g., "experiments help to develop medications") and *consequential* (e.g., "it helps save lives") reasons. These types of reasons are inferences drawn from background knowledge, personal experiences, and the claims of others. However, in science high-quality arguments tend to have an empirical base. In other words, scientific claims need to be supported by genuine evidence and they must have sufficient evidentiary support to be considered acceptable or valid. As a result, this type of framework does not emphasize the ways in which scientific arguments differ from arguments in other fields (e.g., history or theology) or between people in everyday contexts.

The framework of Schwarz et al., as with standard applications of Toulmin's framework, can provide some focus and leverage in terms of issues of content. For example, although the framework does not assess the conceptual quality of an assertion or the reasons used to support the assertion directly, the content of the argument is examined to make judgments about the acceptability or the relevance of the reasons. This requires a researcher to determine whether an argument, as a whole, makes sense. However, if a researcher is interested in assessing the conceptual quality of an assertion as a measure of conceptual change, an additional analysis must be conducted. In exchange for this trade-off, however, this framework enables researchers to compare the argumentation practices of students across a broad range of topics or contexts.

Domain-Specific Analytic Frameworks

We now turn our attention to four frameworks that are more domain specific in terms of their applicability and the criteria that are used to evaluate the nature of arguments. To illustrate this type of domain-specific analytical approach, we discuss the frameworks developed by Zohar and Nemet (2002), Kelly and Takao (2002), Lawson (2003), and Sandoval (2003; Sandoval and Millwood, 2005). Although other excellent examples of domain-specific frameworks can be found in the literature (e.g., Kuhn & Reiser, 2005; McNeill et al., 2006; Zembal-Saul et al., 2003), we chose these frameworks because they provided a broad range of approaches for discussion.

Zohar and Nemet's Focus on the Content of the Justification in an Argument. Zohar and Nemet's (2002) framework was designed to evaluate the quality of written arguments generated by students based on the content of the justification. Zohar and Nemet define an argument as consisting "of either assertions or conclusions and their justifications; or of reasons or supports" (p. 38) and argue that argumentation is a type of informal reasoning because it "involves reasoning about causes and consequences and about advantages and disadvantages, or pros and cons, of particular propositions or decision alternatives" (p. 38). Given this perspective, Zohar and Nemet (2002) suggest that students need to learn the importance of "grounding decisions on reliable knowledge" (p. 40) and that good arguments "include true, reliable, and multiple justifications" (p. 40).

Mechanics of the Framework. In their framework, strong arguments have multiple justifications to support conclusions that incorporate relevant, specific, and accurate scientific concepts and facts. Weak arguments consist of nonrelevant justifications. Conclusions that do not include some type of justification are not considered arguments. Rather than attempting to characterize the components of a particular justification, Zohar and Nemet describe how students incorporate scientific ideas into their arguments. These categories include (a) no consideration of scientific knowledge, (b) inaccurate scientific knowledge, (c) nonspecific scientific knowledge (e.g., "we need to do more tests before we can reach a conclusion"), or (d) correct scientific knowledge. This type of approach enables Zohar and Nemet to sidestep many of the reliability and validity issues associated with Toulmin's framework by collapsing Toulmin's data, warrants, and backings into a single category. This approach also provides valuable information about the content of arguments that students generate.

TABLE 2
The Sample Argument Coded Using Zohar and Nemet's Analytic Framework

Component of the Argument	Code	Scientific Knowledge
I think. . . all objects in the same surroundings become the same temperature even if an object produces its own heat energy.	Claim	Not coded
This is true because on the lab that we did all the temperatures were in their 20s which proves that the room temperature changes the objects to the same as the room.	Relevant justification	Correct scientific knowledge
Therefore, even though they may feel different, the objects are actually within a few degrees of each other.	Relevant justification	Incorrect scientific knowledge

Application to the Sample Argument. Table 2 shows how Zohar and Nemet's framework would be applied to the sample argument. Rather than classifying the statement "on the lab we did all the temperatures were in their 20s" as data and the statement "which proves that the room temperature changes the objects to the same as the room" as a warrant, Zohar and Nemet's analytic framework treats these comments as a single justification. The statement, "therefore, even though they may feel different, the objects are actually within a few degrees of each other" is also classified as a justification because it is used as a way to support the validity of the claim (instead of as a qualifier as per the Toulmin model). From the perspective of Zohar and Nemet's framework, this argument would be considered strong; the claim is supported by two relevant justifications, one of which includes specific and accurate scientific knowledge. The other justification, although it refers to a specific piece of scientific knowledge, is inaccurate from a scientific perspective.

Synthesis in Terms of Structure, Justification, and Content. Zohar and Nemet's framework focuses most heavily on the issues of justification and content. One of the potential benefits of this approach is that the framework enables researchers to determine how often students use scientific knowledge to support an idea and under what conditions. For example, Zohar and Nemet's work suggests that most students (90% in their study) are able to formulate a simple argument, consisting of a claim with a single relevant justification without any formal training about "what counts" as a good argument in science. However, very few of these students (16%) used correct, specific biological knowledge as part of their justification. However, after explicitly teaching students about argument quality and relevant scientific content, Zohar and Nemet observed an increase in both the quality of students' arguments (in terms of numbers of justification used to support a claim) and the how often they used specific biological knowledge as part of their justification. This finding, which mirrors the findings of Schwarz et al. (2003), suggests that students do not refer to specific scientific content to justify their claims unless they have an adequate conceptual understanding of the subject in question and they have an opportunity to rehearse constructing arguments for themselves. This indicates that content knowledge and argumentation practices are intimately linked.

While Zohar and Nemet's framework offers several affordances in terms of issues relating to content and justification in arguments, the framework involves some limitations as well. First, the content of the claim is not evaluated in this analytic framework. This is not an issue when examining the content of an argument generated in response to a socioscientific dilemma (as in Zohar and Nemet's study); in this context, valid opposing claims can be made from multiple perspectives. However, when arguments are generated to articulate and support an explanation for a natural phenomenon, the content of the claim is important. For example, the claim "all objects in the same surroundings are the same temperature *even if* an object produces its own heat energy," (see Table 2) indicates that this student reached an inaccurate conclusion. Thus, this framework is not suitable for examining content issues regarding the sufficiency, usefulness, or accuracy of a claim, assertion, or explanation.

Another constraint of this analytic framework is that it does not include an assessment of how well a student takes into account *all* available information when generating an argument. Cognitive studies of students' experimentation across scientific and quasi-scientific domains reveal that students often fail to see patterns emerging across experiments and often ignore anomalous data or distort them to match their personal beliefs (Chinn & Brewer, 1998; Zeidler, 1997). As a result, students may construct elaborate arguments consisting of several relevant justifications that include accurate scientific knowledge (which would be considered a strong argument using the Zohar and Nemet analytic framework), but the claim might still involve inaccuracies if the students did not coordinate the claim with all available evidence. This constraint is significant, especially because justifications for scientific claims are often based on interpretations of data gathered across multiple experiments.

Kelly and Takao's Focus on the Epistemic Levels of Propositions in an Argument. Kelly and Takao (2002) and Takao and Kelly (2003) developed their analytic framework to analyze longer and more complex written arguments found within term papers by students in an oceanography course. The term papers required students to support an abstract theoretical conclusion based on multiple data representations. The arguments generated by these students often contained multiple propositions to support their particular explanatory conclusion. Kelly and Takao's analytic framework focuses on the relative epistemic status of these propositions and how these propositions are linked together to form persuasive arguments. Kelly and Takao relied heavily on rhetorical studies of science writing in the development of this approach (e.g., Bazerman, 1988; Latour, 1987).

Mechanics of the Framework. To use this framework, a researcher must first identify the propositions found in an argument and then classify them based on epistemic level. These epistemic levels are defined by discipline-specific constructs and reflect a general distinction between lower-level descriptions of data and epistemologically higher-level appeals to theories within a particular domain. Once classified, the researcher identifies how the propositions are linked together and then uses this information to produce a graphical representation of the structure of an argument. This representation is then used to examine the types of propositions an individual uses in his or her writing and how the author coordinates the various propositions into an argument.

Application to the Sample Argument. Figure 4 applies the framework to the sample argument. The sample argument, although relatively short, represents a high level of structural quality from the perspective of this framework: A theoretical conclusion (Box 1) is explicitly linked to a trend that is observed in the data (Box 2) and then to an explanation of how this trend is relevant to the conclusion (Box 3). The student then describes how the conclusion accounts for an everyday observation (Box 4). This argument, therefore, integrates statements from a variety of epistemic levels, indicating higher structural quality according to the framework.

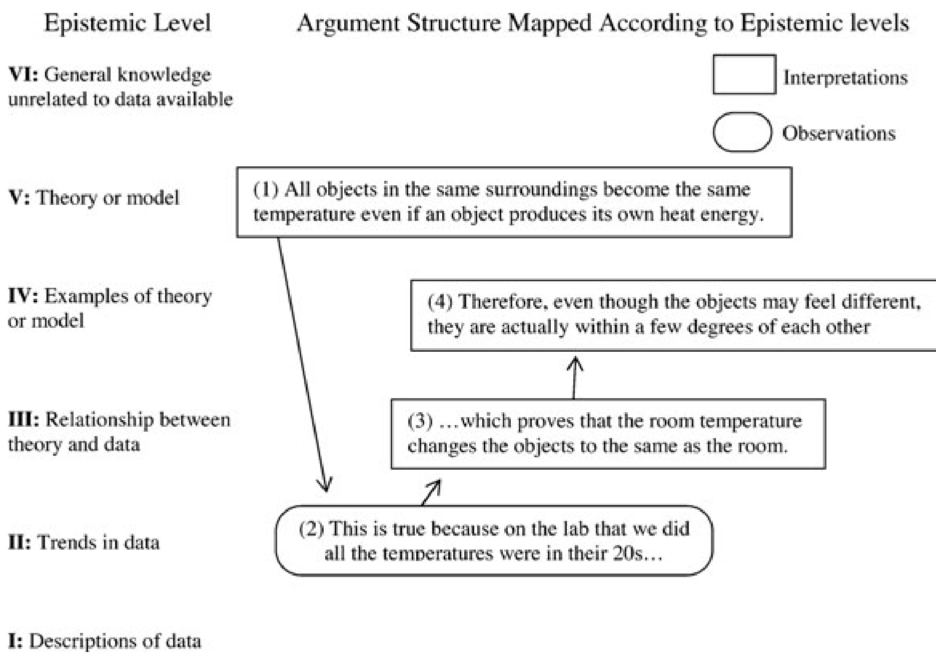

Figure 4. Argument structure of the sample argument based on the epistemic levels of argument framework developed by Kelly and Takao (2002). Comments that reflect observations are depicted as circles and interpretations as squares.

Synthesis in Terms of Structure, Justification, and Content. Kelly and Takao's framework offers substantial and novel affordances in terms of issues relating to structure and justification. The examination of the epistemic statuses of the various propositions incorporated within an argument provides a way to characterize the types of propositions students use to support their conclusions and the extent to which individuals rely on one type of proposition over another. It also provides a way to examine and assess the structure of longer and more complex arguments. For example, the term papers examined by Kelly and Takao required students to formulate a line of evidence based on multiple data representations, reference-specific data, identify specific geological features, and then explain how these features related to their abstract theoretical conclusions. Using Toulmin's argument pattern to analyze the structure of these extended arguments would have neglected critical aspects of argument, such as the importance of moving from specific grounded claims to more generalized statements or the importance of embedding claims within a larger argument (Bazerman, 1988; Latour, 1987). Taken as a whole, this research underscores the fact that the nature of the task will influence the type of argument developed by a student and consequently the analytic frameworks needed to assess it.

One limitation of the framework, however, involves another issue of content. This framework lacks an appraisal of (a) the sensibility of the links between propositions and (b) the scientific accuracy of the propositions. The absence of these evaluations makes it difficult to determine whether students understand the theories or how well the data support the conclusions. Kelly and Takao pointed out this limitation in their own analysis. In fact, Kelly and Takao found several discrepancies between how they rated arguments and how the instructor of the oceanography course rated them. This variation could be attributed to the accuracy of the propositions, the appropriateness of linkages, or it could be the result of a lack of sufficient evidentiary support for the students' conclusions (all of which

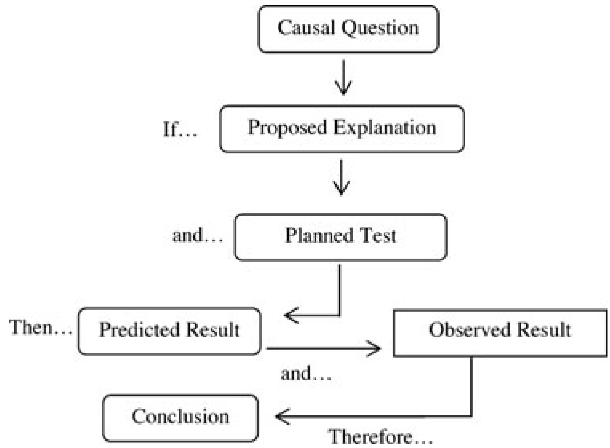

Figure 5. The elements of a hypothetico-predictive argument according to Lawson (2003).

are field-dependent features of an argument). These challenges illustrate the trade-offs researchers face in choosing the aspects of arguments to focus upon in analysis.

Lawson's Focus on the Hypothetical-Deductive Validity of an Argument. Lawson (2003) argues that science educators need to focus their efforts on helping students learn how to generate the type of arguments that are used and valued by scientists rather than focusing on a more general account of argument structure. From his perspective, the goal of developing an argument in science is to "discover which of two or more proposed alternative explanations (claims) for a puzzling observation is correct *and* which of the alternatives are incorrect" (p. 1389, parentheses and emphasis in original). This process requires the generation of an argument that not only presents a tentative explanation that may be correct but also outlines tests based on the generation of specific predictions and the analysis of evidence. Lawson describes this type of argument as a *hypothetico-predictive* argument. According to Lawson, this type of argument, which evaluates the validity of alternative explanations based on hypothetico-deductive reasoning, is much more convincing than arguments that rely on evidence, warrants, and backings to convince others of the validity of a claim because it can provide evidence for one explanation and at the same time provide evidence against others.

Mechanics of the Framework. Figure 5 illustrates the elements of a hypothetico-predictive argument as outlined by Lawson. As shown, the process begins with a perplexing observation that provokes a causal question and the generation of one or more tentative explanations. Once generated, these explanations must be tested to establish their validity. To test the validity of an explanation, one must begin by assuming that the explanation is correct. Next, one must imagine a test that together with the explanation should produce one or more specific observable results. The words "if/and/then" link the explanation and the imagined test to the prediction. Once a test is planned and conducted, the observed results constitute evidence. This evidence is then compared with the prediction. The match or mismatch of the evidence and prediction then leads to a conclusion regarding the validity of the explanation. Lawson indicates that evaluations of overall quality of this type of argument should focus on deductive validity rather than the presence and strength of warrants, which he contends, is the same criterion used by scientists to assess the quality of arguments generated by the scientific community.

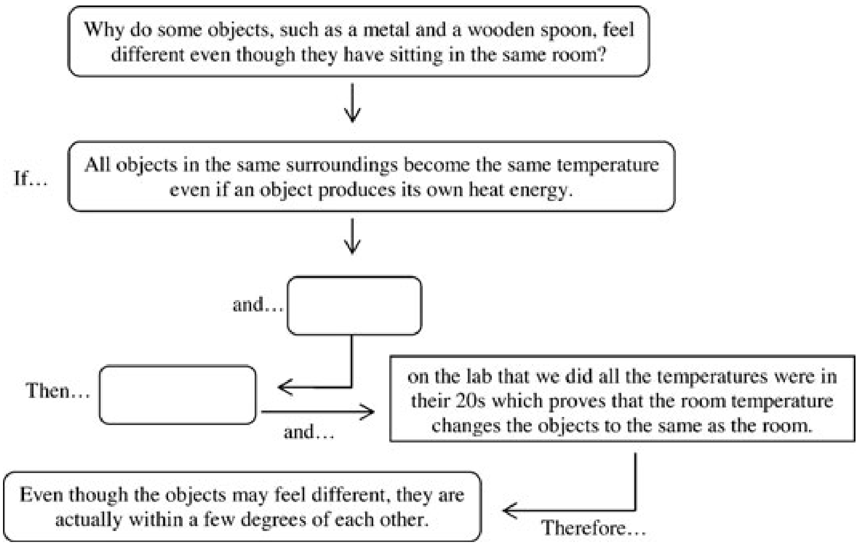

Figure 6. The sample argument assessed using Lawson's (2003) hypothetico-predictive framework.

Application to the Sample Argument. The main objective of a hypothetico-predictive argument is to establish the validity of an explanation. Accordingly, if the sample argument represents high quality from a scientific perspective, we should be able to organize the constituent statements into the form of a hypothetico-predictive argument. As shown in Figure 6, the sample argument fails this test because the student did not describe tests for the validity of their explanation or the predicted results for those tests. Although the sample argument is strong from the structural perspective advocated by informal logicians, Lawson's framework highlights the difficulties that students have generating an argument using these domain-specific criteria. As with many arguments generated by students (e.g., Kuhn & Reiser, 2005), this argument focuses on the articulation and clarification of the author's interpretation of the phenomenon rather than attempting to test the validity of the explanation by coordinating claim and evidence or by discounting alternative explanations for the same phenomenon.

Synthesis in Terms of Structure, Justification, and Content. Lawson's framework obviously focuses most heavily on the issues of justification. While the form of justification required by Lawson's framework is not applicable across all contexts in science, it does parallel certain discipline-specific applications and contexts. Lawson's work (Lawson, 1985; Lawson, Drake, Johnson, Kwon, & Scarpone, 2000) suggests that students often have difficulty developing hypothetico-predictive arguments because they have not developed the reasoning needed to generate and test hypotheses. As a result, when students are specifically asked to generate an argument in this form, they tend to produce arguments with missing or confused elements and their predictions often do not follow from their explanations or planned tests.

While structure is not conceived of in the same manner as in the other framework discussed here, Lawson's framework clearly does focus on issues of structure in a very precise manner. Only arguments that match the complete template of components outlined in Figure 5 are considered strong arguments from the perspective of this framework. This template is very specific in terms of the scientific disciplines and contexts to which it is applicable. It fits a traditional empirical model of hypothesis testing and therefore might apply less well, for example, in terms of science conducted with archival data sets or observational contexts such as certain subfields of geology. As a result, the framework is very specific in terms of the scientific disciplines and contexts to which it applies, but for these disciplines and contexts it provides a strong structural model to guide instruction and student reasoning.

In terms of issues of content, as with many frameworks, Lawson's framework assumes that the content of the various components of the student's argument should cohere and be relevant. Lawson's framework does not, however, include an explicit rubric or template to assess content and conceptual relevance.

Sandoval's Focus on Conceptual and Epistemological Aspects. The sixth and final framework, which was developed by Sandoval (2003) and Sandoval and Millwood (2005), assesses students' arguments in terms of field-dependent criteria including conceptual and epistemological quality. Sandoval argues that scientific arguments, and the methods used to assess them, should reflect the "often tacit epistemological commitments" (Sandoval, 2003, p. 8) established and shared by the individuals who participate in a particular discipline about "the kinds of questions worth asking, the kinds of answers worth having, and acceptable methods for making them" (p. 8). As a result, this approach is the most discipline and subject matter specific of the frameworks that we examine.

Mechanics of the Framework. According to Sandoval's framework, the conceptual quality of an argument measures how well the student has (a) articulated causal claims within a domain-specific theoretical framework and (b) warranted these claims using available data (see Table 3). The epistemological quality of an argument examines whether or not a student has (a) cited sufficient data in warranting a claim, (b) written a coherent causal explanation for the phenomenon (see Sandoval, 2003), and (c) incorporated appropriate rhetorical references when referencing data (see Sandoval & Millwood, 2005).

To use this framework for a specific topic, a researcher must identify the causal elements that are necessary to construct a complete explanation for the phenomenon under investigation and the type of data needed to warrant each element. In the context of the sample argument, for example, a conceptually complete explanation should (a) clarify how heat transfers through different objects at different rates, (b) describe how objects become the same temperature as their surroundings, and (c) illustrate how objects can feel different and still be the same temperature because of differences in conductivity. To provide sufficient data to warrant this type of explanation, students must present data that (a) contrast the rate of heat transfer in different objects, (b) compare the temperatures of various objects, and (c) document how these objects feel.

Application to the Sample Argument. Although the sample argument appears to be of relatively high quality from the perspectives of many of the frameworks discussed thus far, Sandoval's framework highlights substantial conceptual inadequacies (see Figure 7). First, the sample argument does not articulate all of the causal elements needed to explain why objects feel different. Although the student discusses thermal equilibrium and sensation, the student neglects to discuss heat transfer or to provide a reason for differences in thermal sensation. As a result, the student is unable to provide a causal mechanism that can explain *why* objects feel different. Furthermore, this student does not discuss thermal conductivity or the differences between heat and temperature. In terms of warrants, this framework suggests that the student did not include critical data to support all of his or her ideas. The claim, "all objects in the same surroundings become the same temperature even if an object produces its own heat energy," is explicitly linked to a trend that is observed in the data

TABLE 3
The Framework of Sandoval (2003) and Sandoval and Millwood (2005)

	Conceptual Quality		Epistemological Quality		
Causal Element	Articulation of Claim	Data Included to Support Claim	Sufficiency of Cited Data	Causal Coherence	Rhetorical Reference
Environmental pressure	Describes change in some environmental factor that could cause differential survival	Data showing change in a factor over time	Level 4: Full Level 3: Key Level 2: Some Level 1: None	Level 4: Strong Level 3: Good Level 2: Poor Level 1: None	Level 5: Interpretation Level 4: Assertion Level 3: Description Level 2: Pointer Level 1: Inclusion
Individual effect	Explains how environmental change effects individuals	Data justifying link between environmental change and claimed effect	Level 4: Full Level 3: Key Level 2: Some Level 1: None	Level 4: Strong Level 3: Good Level 2: Poor Level 1: None	Level 5: Interpretation Level 4: Assertion Level 3: Description Level 2: Pointer Level 1: Inclusion
Differential trait	Identification of a trait that distinguishes survivors from nonsurvivors	Data that compare this trait in live and dead individuals during the affected time period	Level 4: Full Level 3: Key Level 2: Some Level 1: None	Level 4: Strong Level 3: Good Level 2: Poor Level 1: None	Level 5: Interpretation Level 4: Assertion Level 3: Description Level 2: Pointer Level 1: Inclusion
Selective advantage	Provides a mechanism for the advantage gained by individuals who posses the trait	Provides behavioral data that links the trait to a function that only the individuals with that trait can perform	Level 4: Full Level 3: Key Level 2: Some Level 1: None	Level 4: Strong Level 3: Good Level 2: Poor Level 1: None	Level 5: Interpretation Level 4: Assertion Level 3: Description Level 2: Pointer Level 1: Inclusion

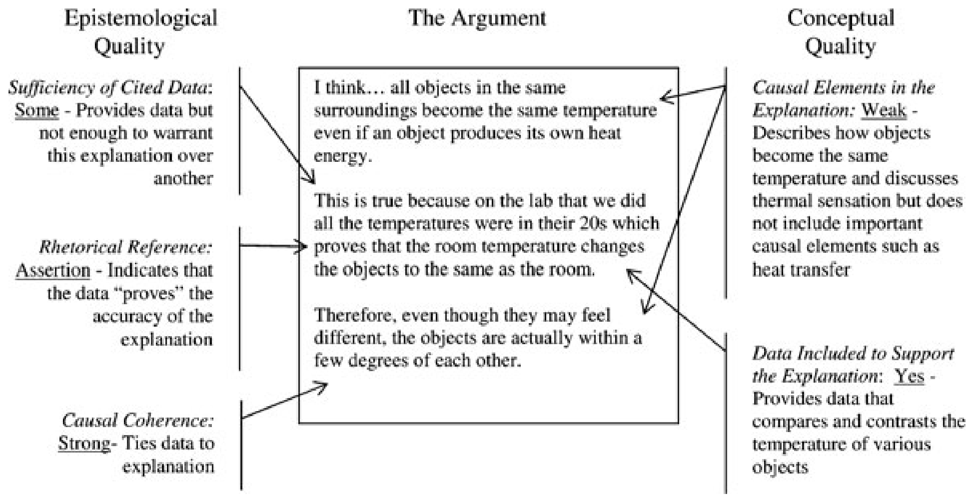

Figure 7. The sample argument assessed using the framework of Sandoval (2003) and Sandoval and Millwood (2005).

and an explanation of how this trend supports the claim. However, the student neglects to provide any data supporting the idea that objects can feel different even though they are the same temperature.

In terms of epistemological quality, this argument is also poor. Although the argument is coherent, the student does not provide sufficient data to warrant the claim in comparison to any other claim. Rather than providing observational data about how different objects feel and then linking these data to the actual temperature of these objects, or comparing and contrasting the temperature of objects that produce heat energy to objects that do not, this student relies on a single piece of evidence to support his or her claim. Furthermore, the rhetorical reference used to link the claim and the data is a simple assertion. The comment, "on the lab that we did all the temperatures were in their 20s which proves that the room temperature changes the objects to the same as the room" indicates that this student believes that the results of an experiment are enough to prove an idea right or wrong rather than viewing an experiment as a way to test an idea and to generate evidence that can be used to support a claim. According to Sandoval, these types of rhetorical references are common in arguments constructed by students, while rhetorical references that make a "detailed interpretation of data" in a way that explains how data support a particular claim are more common in professional science texts.

Synthesis in Terms of Structure, Justification, and Content. Sandoval's framework focuses specifically on issues of justification and content and provides less explicit focus on issues of structure. In terms of justification, Sandoval's framework provides information about the types of epistemological criteria that students use to articulate and validate an argument and how these criteria align with the criteria used within particular scientific disciplines. Sandoval argues for the importance of studying rhetorical references because the manner in which students incorporate and refer to data in their writing may reflect their implicit epistemological commitments about the nature and role of data in the generation and evaluation of scientific knowledge. For example, Sandoval and Millwood's (2005) analysis of the arguments constructed by high school students indicates that many students are able to apply their understanding of natural selection to generate an argument that can be used to explain a natural phenomenon that is consistent with the major tenets of natural selection.

However, their research also suggests that students tend to rely on a single piece of data when supporting a particular claim even though they apparently understand the importance of linking claims and evidence (and apparently assume that a single piece of data is enough to prove the veracity of a claim). Students, as a result, often do not include comparisons of data from multiple sources when warranting claims even when such comparisons are needed.

In terms of content, Sandoval's framework offers particular strengths in terms of determining whether students can generate arguments that explain particular phenomena using specific theories, such as natural selection, and warrant these claims using the available data. Obviously, the subject-matter-specific nature of this framework requires significant adaptation for application to other contexts or comparisons across contexts, but that same specificity affords dramatic insights into the conceptual quality of students' ideas and arguments. In this way, Sandoval's framework moves beyond assessing content in terms of whether or not the argument contains normative information, or even whether or not the argument contains relevant information. Sandoval's framework instead assesses content in terms of the overarching explanatory power of the argument within the specific disciplinary context in which the argument is set.

This type of domain- and subject-specific framework can also be used as a way to help students understand the nature and development of scientific knowledge and how to generate and evaluate arguments in a more productive manner. For example, as individuals generate an argument, or critique an argument proposed by another, students must explicitly consider the purposes of an explanation, how well it explains or describes the phenomenon in question, and how well it is supported by available evidence (Sandoval, 2003; Sandoval & Reiser, 2004). This type of focus can help students understand the "practices of the scientific community with its particular purposes, ways of seeing, and ways of supporting its knowledge claims" (Driver, Asoko, Leach, Mortimer, & Scott, 1994, p. 8) and illustrate how science is unique from other ways of knowing.

CONCLUSIONS AND FUTURE DIRECTIONS

Researchers have relied on a broad range of analytical perspectives to examine the nature and quality of student-generated arguments in the context of science education. In this article, we have examined the constraints and affordances of several frameworks designed for this purpose. We will now use these frameworks to provide a synthesis of the ways researchers in science education assess argument by focusing on issues related to structure, content, and justification.

Structure Across the Frameworks

Many of the frameworks included in this review describe the structure of an argument in terms of claims and justification. These "claim" components are fairly straightforward and similar in most of the frameworks. Some refer specifically to *claims* (e.g., Toulmin), whereas others talk about *assertions* (e.g., Schwarz et al.) or *explanations* (e.g., Lawson and Sandoval) but most of the frameworks include a structural claim-like component that requires justification or substantiation. In many of these frameworks (e.g., Toulmin, Zohar and Nemet, Lawson), the claim-like component is viewed as single entity that corresponds roughly to a single proposition (e.g., "all objects in the same surroundings become the same temperature" or "they should not have children"); however, in some (e.g., Sandoval) the claim-like component is more complex and includes several subcomponents that must all be present in order for the claim-like component (or explanation) to be considered sufficient.

The frameworks also diverge with regard to the numbers and types of structural components involved in justifying the claim-like component. Toulmin created the standard *data*, *warrants*, *backings*, and *qualifier* justification components that have influenced many subsequent frameworks. For example, Sandoval's framework includes structural components for justification in terms of *data* and *rhetorical references* that link the data to a component of the *explanation*. Some of the frameworks, however, take a minimalist approach to the specification of these justification components and simply include a generic *reason* or *justification* component. Some researchers make this choice to simplify the difficulties that have been encountered by researchers in reliably differentiating between the *data*, *warrants*, and *backings* distinctions in Toulmin's framework. Other researchers (e.g., Zohar and Nemet and Schwarz et al.) choose a simpler generic structural component for justification because their framework focuses more on the nature or quality of the justification rather than the absence or presence of a structural component devoted to justification.

Other frameworks attempt to examine the complexity of an argument by using additional structural components that build on this claim/justification dichotomy. For example, Lawson's framework takes a structurally more complex and specific approach with regard to the components of an argument. The structural components for an argument in this framework include a *planned test* for the *proposed explanation* (Lawson's claim-like component) along with *predicted results* for that test if the *proposed explanation* is valid. The *predicted results* must then be compared with *observed results* of the *planned test* to determine whether or not the *proposed explanation* is supported. Lawson's framework is, therefore, very specific about the required combination of structural components that students must incorporate in their arguments. This type of specificity is useful in instructional contexts because it provides an explicit structural model to guide student work and reasoning but also makes the framework very domain- and context-specific in terms of scientific argumentation practices (i.e., this template would not necessarily fit with various forms of argumentation considered valid in fields of science that do not rely on hypothetico-deductive reasoning to validate explanations).

Kelly and Takao's framework is exceptional among the frameworks reviewed here because it does not make a fundamental structural distinction in terms of the claim/justification dichotomy. Kelly and Takao's framework instead focuses on larger pieces of reasoned discourse where a proposition might be considered a justification with respect to a second proposition but be considered a claim with respect to a third proposition. As a result, Kelly and Takao's framework distinguishes between structural components in terms of the epistemic abstractness of each proposition and the connections between the individual propositions. The structural units of analysis in Kelly and Takao's framework are, therefore, propositions and their connections. This represents a significantly different way of thinking about structure.

This diversity of perspectives has provided considerable insight into the ways students generate arguments in the context of science education. For example, research that examines the structure of argument in terms of the claim/justification dichotomy has demonstrated that students often produce one-sided arguments (Sadler, 2006; Schwarz et al., 2003; Schwarz & Glassner, 2003) and do not attempt to support their claims with multiple justifications (Sandoval & Millwood, 2005; Zohar & Nemet, 2002) or attempt to show why one claim is more acceptable than another (Bell & Linn, 2000; Kuhn & Reiser, 2005). On the other hand, the research that has focused on claim-like components indicates that students struggle to generate complex and sufficient explanations for the phenomenon under investigation (McNeill et al., 2006; Sandoval, 2003; Sandoval & Millwood, 2005). Finally, research that has examined the overall structure of an argument indicates that students often struggle with coherence and linking ideas together in appropriate or meaningful ways

(Kelly & Bazerman, 2003; Kelly, Regev, & Prothero, 2005; Kelly & Takao, 2002; Lawson, 2002).

Future Directions. Much research on argumentation in science education has focused on argument structure as the primary criterion for quality. Furthermore, much of this research has relied on the claim/justification distinction to define structure. More recently, approaches such as Kelly and Takao's have begun to think about structure in a different manner. In addition, other frameworks have begun to link issues of structure to those of content and justification so that assessments of quality are not based solely on the frequency counts of the various structural components. Approaches that emphasize structure clearly offer fantastic affordances in terms of providing templates for instruction and analysis that can be applied across wide ranges of contexts. Structural frameworks will, therefore, continue to play critical roles in science education research. However, future research will also need to examine the connection between structural components and the relevance, sufficiency, and accuracy of their content as well as the epistemic nature of inherent justification strategies.

Content Across the Frameworks

While most of the authors of the reviewed frameworks would agree that content is important, only the frameworks developed by Sandoval and Zohar and Nemet include specific rubrics or hierarchies for assessing content quality within arguments. For example, although Toulmin's argument pattern and Schwarz and colleagues' framework consider the content of an argument to make judgments about the acceptability or the relevance of reasons, these frameworks do not directly assess the conceptual quality of an assertion or the reasons used to support assertions. Similarly, while Lawson's framework assumes that the content should cohere, his framework does not include an explicit rubric or template to assess content and conceptual relevance.

Kelly and Takao's framework, on the other hand, represents and interesting intermediate focus on content. While Kelly and Takao do not measure content quality directly, the epistemic levels of their framework are defined by discipline-specific constructs and reflect a general distinction between lower-level descriptions of data and epistemologically higher-level appeals to theories within a particular domain. Therefore, while the framework lacks an appraisal of the sensibility of the links between propositions or the scientific accuracy of the propositions, issues of content are intimately integrated within the framework.

Zohar and Nemet focus more directly on issues of content. Rather than attempting to characterize the components of a particular justification, Zohar and Nemet chose to examine how often students incorporate scientific ideas into their arguments as a way to justify their claims. The framework does not, however, assess the content of the claim. This is not an issue when examining the content of an argument generated in response to a socioscientific dilemma in which valid opposing claims can be made from multiple perspectives (as in Zohar and Nemet's study). However, when arguments are generated to articulate and support an explanation for a natural phenomenon, the content of the claim increases in importance. Another constraint of this analytic framework in terms of content is that it does not include an assessment of how well a student takes into account *all* available information when generating an argument.

Of the frameworks reviewed here, Sandoval's framework offers the highest mechanical specificity in terms of content quality. This framework examines content issues by assessing how well an argument explains a particular phenomenon using a specific theory, such as natural selection, and warrants the various elements of this explanation using available data. Obviously, the subject-matter-specific nature of this framework requires significant adaptation for application to other contexts or comparisons across contexts, but that same specificity affords dramatic insights into the conceptual quality of students' ideas and the substance of their arguments. In this way, Sandoval's framework moves beyond assessing content in terms of whether or not the argument contains normative information, or even whether or not the argument contains relevant information. Sandoval's framework instead assesses content in terms of the overarching explanatory power of the argument within the specific disciplinary context in which the argument is set.

The diversity of perspectives about content and the specific criteria that have been incorporated into some of these frameworks have provided valuable insights about the substance of student-generated arguments. For example, some researchers have found that students often use appropriate and relevant reasons to support inaccurate claims (Clark & Sampson, 2006b), use irrelevant forms of justification to support a claim (Kuhn & Reiser, 2005; McNeill et al., 2006; Schwarz et al., 2003), or include insufficient amounts of justification in terms of content (Sandoval, 2003; Sandoval & Millwood, 2005). Research examining the nature of the explanations that students attempt to justify suggests that students often do not focus on key issues content issues such as causality; instead students tend to provide explanations that are simple descriptions of observations (Kuhn & Reiser, 2005; Sandoval & Millwood, 2005; Sandoval & Reiser, 2004; Zembal-Saul et al., 2003). Overall, it seems that student-generated arguments often lack substance, leave out important points, or contain components that are inaccurate and/or irrelevant in terms of content.

Future Directions. Much of the research that has examined the content of arguments generated by students in the context of science education has focused on the content of the justifications or how well various components of an argument fit together as a whole rather than focusing specifically on the content of the claim/assertion/explanation. This is due, in large part, to the relatively high proportion of studies that have examined how students generate arguments in response to socioscientific dilemmas (such as asking students to generate an argument for or against the use of animals in medical research). As researchers in our field move into new areas of research focusing on argument and conceptual change, however, we will need new tools with more explicit focus on the content, logical coherence, relevance, and explanatory power of the claim/assertion/explanation. Sandoval's framework may provide excellent guidance for this type of research because of the emphasis this approach places on explanatory power and sufficiency of explanations. Further development of these frameworks will give researchers the ability to trace or map the ideas of students in terms of their overall understanding of the topic under investigation (as evidenced by the causal mechanisms and core concepts included in their arguments).

Justification Across the Frameworks

All of the frameworks reviewed here focus on justification in some manner. Perspectives on justification and the criteria used to determine quality, however, differ substantially. Some frameworks describe justification in terms of information components (e.g., "data, warrants, and backings" or "reasons") while other frameworks view justification more in terms of thought processes (e.g., "how propositions of various epistemic levels are linked together" or "hypothetico-deductive reasoning"). The criteria for assessing justification quality in these frameworks often reflect this distinction (although significant overlap exists within and across the frameworks). Frameworks that predominately view justification in terms of information components rely on the absence or presence of various pieces of information (e.g., "there is no data to support this causal element" or "the conclusion is supported by two reasons") or a classification of the characteristics of these components (e.g., "the argument includes a *make sense* reason rather than a *consequential* reason" or "the reason includes accurate and relevant scientific knowledge") as indicators of quality. On the other hand, frameworks that describe justification as a thought process tend to examine how the various components of the argument fit together (e.g., "the match or mismatch of the evidence and prediction" or "moving from specific descriptions of data to more generalized statements") in order to assess quality.

This diversity of perspectives has provided considerable insight about the ways students justify their ideas through argument. For example, research focusing on justification in terms of information components has demonstrated that students tend to rely on insufficient or inappropriate information to justify their data. This work suggests that most students tend to focus on the articulation of a claim, viewpoint, or explanation rather than attempting to support it (Kuhn & Reiser, 2005; McNeill et al., 2006; Sadler, 2006; Zembal-Saul et al., 2003; Zohar & Nemet, 2002). When students do include a justification in an argument, they tend to rely on inferences, personal experiences, and authority figures (Kelly & Chen, 1999; Kuhn & Reiser, 2005; Schwarz et al., 2003; Zohar & Nemet, 2002). Even when students use appropriate (or desirable) forms of justification, they often do not include enough information to warrant one claim, viewpoint, or explanation over another (Sandoval, 2003; Sandoval & Millwood, 2005; Sandoval & Reiser, 2004). On the other hand, research focusing more on the "thought process" of justification has demonstrated that students tend to rely on different forms of reasoning than those used in science (at least when they are asked to generate an argument to establish the validity or acceptability of a theory, explanation, or model). It seems that many students struggle to generate valid hypothetico-predictive arguments (Lawson, 2002, 2003), to move from concrete observations to more abstract ideas in an argument (Kelly et al., 2005; Kelly & Takao, 2002), or to use appropriate rhetorical references ("this shows" or "this suggests" rather than "this proves") when attempting to establish the validity or acceptability of a claim, explanation, or viewpoint (Sandoval & Millwood, 2005).

Future Directions. Our field now needs to focus our attention on developing a better understanding of why students engage in justification the way they do. While some work has been conducted in this area (e.g., Brem & Rips, 2000), many questions remain. Why do students use insufficient or inappropriate information to justify their ideas? Why do students struggle to produce arguments with coherent justification, appropriate forms of reasoning, and suitable rhetorical moves? Do these struggles result from a lack of familiarity with the topic or a lack of understanding of the cognitive, epistemic, and social aspects of argument generation (e.g., Duschl, 2008) or both? Similarly, we need to better understand the criteria that students use to determine what evidence is most persuasive or to warrant one idea over another. Answers to these types of questions will enable science education researchers to develop better curricular materials, instructional practices, and technology-enhanced learning environments to promote and support more productive argumentation inside the classroom. In order to engage in these investigations, however, we will need frameworks that analyze justification more broadly and authentically in terms of epistemically appropriate thought processes connecting relevant information components into coherent explanations. As with structure and content, our field has developed frameworks that tend to atomize arguments so that we see the proverbial trees, but not the proverbial forest composed of those trees. In other words, we now need frameworks that allow us to analyze the overarching patterns of justification as related both to the content and to the structure of arguments.

Final Thoughts

Empirical and theoretical research in science education has illustrated the important role that argument generation and evaluation play in science and science education (Lemke, 1990; Osborne, 2002). The analysis of students' arguments, such as the sample argument in this review, can provide a great deal of information about students' understanding of scientific content (e.g., the theories, laws, and ideas that are important in science), students' scientific reasoning, students' epistemological commitments (e.g., what counts as warranted knowledge), and students' ability to communicate and justify ideas to others.

In this article, we have examined the constraints and affordances of several frameworks designed to examine these issues by focusing on issues of structure, content, and justification. We now return a question that we posed at the beginning of this review. Did the author of the sample argument produce a high-quality scientific argument that articulates and justifies an explanation for why some objects, such as a metal and a wooden spoon, feel different even though they have sitting in the same room for several hours? The answer, as illustrated by this review, depends on the framework chosen for the analysis.

Some of the frameworks, such as the ones developed by Toulmin (1958), Zohar and Nemet (2002), Schwarz and colleagues (2003), and Kelly and Takao (2002) suggest that the student generated an appropriate scientific argument. The frameworks developed by Sandoval (2003) and Lawson (2003), however, suggest that the sample argument is inadequate from a scientific perspective. Furthermore, the four frameworks that assess the sample argument favorably do so for different reasons and the two frameworks that assess it unfavorably also do so for different reasons. These differing assessments result from both the divergent foci of the frameworks in terms of relative weights placed on structure, content, and justification as well as differences in how the frameworks define structure, content, and justification. We have considered these differences in detail in the discussion of the individual frameworks as well as in the syntheses above.

The diversity of perspectives on argument assessment represented by these frameworks suggests several overarching messages for consideration. First, it is important for researchers to understand that analytic frameworks, such as those reviewed here, are tools created for specific tasks to investigate specific questions. Frameworks, therefore, are not fully interchangeable, and the foci of each framework require consideration before comparing the results of various studies. Another important message, related to the first, underscores how much information readers need to interpret the results of a study; it is simply not enough to say that a given intervention supports students in creating "high"- or "low"-quality arguments. An audience needs very specific details about the nature of the analytic foci as well the underlying assumptions about "what counts" as quality to interpret findings. Explicit sharing of these details among researchers will improve communication and comparison of results across studies. The creation of a shared theoretical foundation and taxonomy for planning, executing, and communicating the findings of future research on argument and argumentation in science education would facilitate this collaboration and communication. The basic taxonomy of structure, content, and justification employed by this review could provide a foundation upon which to develop a more detailed taxonomy for these purposes.

This review also suggests a number of overarching messages regarding the nature of research in our field. First, this review suggests that much research to date has focused on very atomized aspects of students' arguments. While this atomized approach has proven fruitful, future research will hopefully include more holistic considerations of the quality of the arguments that students articulate. This work, however, will require new approaches that examine the structural, conceptual, epistemic, and social aspects of argument generation in a more synergistic fashion rather than looking at each of these aspects independently. Second, this review suggests that much research on argument in science education has thus far focused on the identification of patterns and themes in students' arguments (e.g., "students tend to produce arguments that lack sufficient justification" or "students tend to produce arguments that have a simplistic structure") rather than focusing on the underlying reasons for these patterns. While research to date has provided substantial and valuable empirical foundations for science educators to build upon, more research examining the underlying causes of these patterns and themes will prove valuable in developing new curricular materials, instructional approaches, and technology-enhanced learning environments to promote and support more productive argumentation inside the classroom.

REFERENCES

Abell, S. K., Anderson, G., & Chezem, J. (2000). Science as argument and explanation: Exploring concepts of sound in third grade. In J. Minstrell & E. H. Van Zee (Eds.), Inquiry into inquiry learning and teaching in science (pp. 100–119). Washington, DC: American Association for the Advancement of Science.

Bazerman, C. (1988). Shaping written knowledge: The genre and activity of the experimental article in science. Madison: University of Wisconsin Press.

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. Linn, E. A. Davis, & P. Bell (Eds.), Internet environments for science education. Mahwah, NJ: Erlbaum.

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. International Journal of Science Education, 22(8), 797–818.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. Cognitive Science, 24(4), 573–604.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. Journal of Research in Science Teaching, 35(6), 623–654.

Cho, K.-L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. Educational Technology Research and Development, 50(3), 5–22.

Clark, D., & Sampson, V. (2006a). Personally-seeded discussions to scaffold online argumentation. International Journal of Science Education, 29(3), 253–277.

Clark, D., & Sampson, V. (2006b). The quality of argumentation supported by personally-seeded discussions. In T. Koschmann, T.-W. Chan, & D. Suthers (Eds.), Computer supported collaborative learning 2005. Mahwah, NJ: Erlbaum.

Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. Educational Psychology Review, 19(3), 343–374.

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(3), 287–313.

Duschl, R. (2000). Making the nature of science explicit. In R. Millar, J. Leach, & J. Osborne (Eds.), Improving science education: The contribution of research. Philadelphia: Open University Press.

Duschl, R. (2008). Quality argumentation and epistemic criteria. In S. Erduran & M. Jimenez-Aleixandre (Eds.), Argumentation in science education: Perspectives from classroom-based research (pp. 159–175). Dordrecht, the Netherlands: Springer.

Duschl, R. A., Ellenbogan, E., & Erduran, S. (1999). Promoting argumentation in middle school classrooms: A project SEPIA evaluation. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Boston, MA.

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

Eichinger, D., Anderson, C. W., Palincsar, A. S., & David, Y. M. (1991). An illustration of the roles of content knowledge, scientific argument, and social norm in collaborative problem solving. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 1991.

Erduran, S. (2008). Methodological foundations in the study of argumentation in science classrooms. In S. Erduran & M. Jimenez-Aleixandre (Eds.), Argumentation in science education: Perspectives from classroom-based research. (pp. 47–69). Dordrecht, the Netherlands: Springer.

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

Erkens, G., Kanselaar, G., Prangsma, M., & Jaspers, J. (2003). Computer support for collaborative and argumentative writing. In E. D. Corte, L. Verschaffel, N. Entwistle, & J. v. Merriënboer (Eds.), Powerful learning environments: Unraveling basic components and dimensions (pp. 157–176). Amsterdam: Pergamon and Elsevier Science.

Forman, E. A., Larreamendy-Joerns, J., Stein, M. K., & Brown, C. A. (1998). "You're going to want to find out which and prove it": Collective argumentation in a mathematics classroom. Learning and Instruction, 8(6), 527–548.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. Journal of Research in Science Teaching, 38(6), 663–687.

Jimenez-Aleixandre, M., Rodriguez, M., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. Science Education, 84(6), 757–792.

Kelly, G. J., & Bazerman, C. (2003). How students argue scientific claims: A rhetorical-semantic analysis. Applied Linguistics, 24(1), 28–55.

Kelly, G. J., & Chen, C. (1999). The sound of music: Constructing science as a sociocultural practice through oral and written discourse. Journal of Research in Science Teaching, 36(8), 883–915.

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. International Journal of Science Education, 20(7), 849–871.

Kelly, G. J., Regev, J., & Prothero, W. (2005). Assessing lines of evidence with argumentation analysis. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX, April 4–7, 2005.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: an analysis of university oceanography students' use of evidence in writing. Science Education, 86(3), 314–342.

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb & H. Bauersfeld (Eds.), Emergence of mathematical meaning. Hillsdale, NJ: Erlbaum.

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

Kuhn, D., & Udell, W. (2003). The development of argument skills. Child Development, 74(5), 1245–1260.

Kuhn, L., & Reiser, B. (2005). Students constructing and defending evidence-based scientific explanations. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

Kuhn, L., & Reiser, B. (2006). Structuring activities to foster argumentative discourse. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Kuhn, T. S. (1970). The structure of scientific revolutions (2nd ed.). Chicago: University of Chicago Press.

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Cambridge, MA: Harvard University Press.

Lawson, A. (1985). A review of research on formal reasoning and science teaching. Journal of Research in Science Teaching, 22(7), 569–618.

Lawson, A. (2002). Sound and faulty arguments generated by preservice biology teachers when testing hypotheses involving unobservable entities. Journal of Research in Science Teaching, 39(3), 237–252.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. International Journal of Science Education, 25(11), 1387–1408.

Lawson, A., Drake, N., Johnson, J., Kwon, Y., & Scarpone, C. (2000). How good are students at testing alternative hypotheses involving unseen entities? The American Biology Teacher, 64(4), 546–553.

Lemke, J. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. Journal of the Learning Sciences, 15(2), 153–191.

Means, L. M., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. Cognition and Instruction, 14(2), 139–178.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21(5), 553–576.

Osborne, J. (2002). Science without literacy: A ship without a sail? Cambridge Journal of Education, 32, 203–215.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. Journal of Research in Science Teaching, 41(10), 994–1020.

Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. Journal of Research in Science teaching, 39(3), 185–204.

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S.-Y. (2001). Influence of oral discussion on written argument. Discourse Processes, 32(2/3), 155–175.

Sadler, T. (2006). Promoting discourse and argumentation in science teacher education. Journal of Science Teacher Education, 17, 323–346.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. Journal of the Learning Sciences, 12(1), 5–51.

Sandoval, W. A., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23(1), 23–55.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education, 88(3), 345–372.

Schwarz, B., B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentative activity. Journal of the Learning Sciences, 12(2), 219–256.

Schwarz, B., & Glassner, A. (2003). The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 227–260). Dordrecht, the Netherlands: Kluwer.

Siegel, H. (1989). The rationality of science, critical thinking and science education. Synthese, 80(1), 9–42.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. International Journal of Science Education, 28(2/3), 235–260.

Stewart, J., Cartier, J. L., & Passmore, C. (2005). Developing understanding through model-based inquiry. In S. Donovan & J. D. Bransford (Eds.), How students learn science in the classroom. Washington, DC: The National Academies Press.

Takao, A. Y., & Kelly, G. J. (2003). Assessment of evidence in university students' scientific writing. Science & Education, 12(4), 341–363.

Toulmin, S. (1958). The uses of argument. Cambridge, England: Cambridge University Press.

Veerman, A. (2003). Constructive discussions through electronic dialogue. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 117–143). Dordrecht, the Netherlands: Kluwer.

Vellom, R. P., & Anderson, C. W. (1999). Reasoning about data in middle school science. Journal of Research in Science Teaching, 36(2), 179–199.

Weinberger, A. (2003). Scripts for computer-supported collaborative learning. Effects of social and epistemic cooperation scripts on collaborative knowledge construction. Munich, Germany: Ludwig-Maximilians-University.

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. Science Education, 81, 483–496.

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2003). Scaffolding preservice teachers' evidence-based arguments during an investigation of natural selection. Research in Science Education, 32, 437–463.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching, 39(1), 35–62.