Assessment of Argument in Science Education: A Critical Review of the Literature

Victor D. Sampson, Douglas B. Clark, Arizona State University
College of Education, Division of Curriculum and Instruction, 203 Payne Hall, Tempe, AZ 85287-0911, USA
Email: victor.sampson@asu.edu, douglas.b.clark@asu.edu

Abstract: Theoretical and empirical research on argumentation in science education has intensified over the last two decades. This article reviews and synthesizes analytical methods used to assess and characterize the nature of rhetorical arguments generated by students as part of this research. This synthesis of the current state of argumentation analysis in science education research provides a comprehensive framework for future research on rhetorical argumentation.

Introduction

Research on argumentation in science education has expanded and intensified considerably over the last decade. In this article, we examine the various analytical methods that have been developed to assess rhetorical argumentation within the context of science education. Our overall goals involve (a) illustrating the logic and assumptions that have pervaded research in this field, (b) summarizing the constraints and affordances of these different approaches, and (c) making recommendations for new directions.

Methods to Assess Argument Quality

Many initial efforts to evaluate students' discursive practices in science have focused on the ways in which students develop arguments or how the arguments generated by students compare to arguments developed by scientists. According to Passmore and Stewart (2002), "Scientific inquiry is fundamentally about reducing the world to order; those reductions take the form of explanations" (p. 188). In order to accomplish this task, scientists make observations, identify patterns in data, then develop and test explanations for those patterns. In constructing an argument, scientists attempt to establish the acceptability of the explanations that they have developed. Studies of scientists doing science indicate that this goal is accomplished by coordinating supportive or contradictory evidence with a particular explanatory or descriptive claim for an observed phenomenon (Kuhn, 1970; Latour, 1987). However, in order to be persuasive, this process must be carried out in way that is 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 proposing tests of claims (Hogan & Maglienti, 2001), (b) the need for coherence between theoretical frameworks and observations of phenomena (Passmore & Stewart, 2002), (c) establishing the credibility of evidence (Driver et al., 2000), (d) parsimony (Sandoval & Reiser, 2004), and (e) the importance of basing arguments on reasoning that is logically valid (Zeidler, 1997). Thus, in order to generate a scientific argument, an individual must learn the kinds of claims scientists make, how they advance them; what kinds of evidence is needed to warrant an argument, and how that evidence can be gathered and interpreted given community standards.

How do students, who are not members of the scientific community, generate arguments? What type of reasoning do students use when they construct arguments and how is it different than scientists? Are students able to assimilate the desired practices of argumentation as a result of classroom instruction? To answer these types of questions, researchers must develop tools to evaluate the quality of the arguments generated by students. Our review of the literature indicates that a diversity of approaches for analyzing rhetorical argumentation exists across studies. Substantial variation is observed, for instance, in the perspectives on argument quality, the questions addressed, the analytical methods adopted, and the specific units of analysis chosen. A critical analysis of these different approaches can provide valuable information about the benefits and limitations of the different analytical methods and assumptions that underlie this field of research.

Toulmin's framework: Examining the structural aspects of arguments

Historically, science education researchers have been heavily influenced by Toulmin's perspectives on argumentation. In *The Uses of Argument* (1958), Toulmin made a distinction between the idealized notions of the logical-formal arguments as they are used in mathematics and how arguments are used in linguistic contexts. 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. According to this framework, a claim is the assertion that an individual makes and data are the facts that a person explicitly appeals to as a foundation for their claim. Warrants are used by individuals to justify why data is relevant to the claim. The strength of the warrant is indicated by including 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 scrutinized. Finally, a rebuttal indicates the "circumstances in which the general authority of the warrant would have to be set aside" (p. 101). Toulmin describes the process of constructing a scientific argument primarily as a process of using evidence, 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 these different structural components. Stronger arguments contain more of these different components than weaker arguments.

Analyses of arguments constructed by students using Toulmin's argument framework have primarily examined how students provide warrants for claims, when they do so, and on what basis (e.g., Jimenez-Aleixandre, Rodrigues, & Duschl, 2000; Kelly, Druker, & Chen, 1998; Osborne, Erduran, & Simon, 2004). These studies have provided a great deal of information about the form of student talk or writing in various setting but have provided little information about how well students engage in argument construction in terms of content quality or normativity. As a result, analytic methods that examine argument quality solely from a structural perspective provide little or no information about how (or if) students' conceptual ideas about the subject matter influence how they coordinate theory with evidence as they construct an argument in support of a particular viewpoint.

Another constraint of Toulmin's argument framework is the fact that it provides little information about *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. However, what counts as an appropriate claim, warrant, backing, or datum 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 the 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 part of their argument? Do students attempt to coordinate their claim with all available data or just the data that supports their particular viewpoint? Answers to these types of questions can provide valuable insights into students' reasoning. However, since these criteria for constructing a scientific argument are grounded in certain logical and epistemological commitments specific to science (Hogan & Maglienti, 2001; Sandoval, 2003), different analytical methods are needed to explore these field-dependent features of an argument.

Zohar and Nemet's framework: Examining the structure and content of arguments

Zohar and Nemet (2002) modified Toulmin's argumentation framework based on the work of Means and Voss (1996) to evaluate the quality of written arguments generated by students based on structure and content. Zohar and Nemet define an argument as consisting "of either assertions or conclusions and their justifications; or of reasons or supports" (p. 38). Therefore, strong arguments have multiple justifications to support a conclusion that incorporate relevant, specific, and accurate scientific concepts and facts. Weak arguments consist of individual non-relevant justifications. Conclusions that do not include some type of justification are not considered arguments. Zohar and Nemet also collapsed Toulmin's data, warrants and backings into a single category to sidestep many of the reliability and validity issues associated with Toulmin's framework (see Kelly et al., 1998 for examples). Justifications are analyzed to determine if they include (a) no consideration of scientific knowledge, (b) inaccurate scientific knowledge, (c) non-specific scientific knowledge (we need to do more tests before we can reach a conclusion), or (d) correct scientific knowledge.

While Zohar and Nemet's framework offers several affordances, the framework involves some limitations as well. The accuracy of the claim itself is not evaluated in this analytical framework. As a result, this framework works better when used to analyze arguments generated in the context of socio-scientific issues (such as family planning based on genetic testing) rather than in the context of scientific debates. In response to socioscientific dilemmas that Zohar and Nemet studied, valid opposing claims can be made from multiple perspectives. However, when arguments are scientific, claims are explanatory conclusions or descriptive frameworks. The accuracy of these revel student's understanding of scientific content and how well they were able to coordinate claims with evidence. Another constraint of this analytical framework is that it does not analyze how well students were able to coordinate claims with all available evidence. Cognitive studies of students' experimentation across scientific and quasiscientific domains reveal that most students have a great deal of difficulty coordinating data with causal ideas or hypotheses (Zeidler, 1997). Students often fail to see patterns emerging across experiments and often ignore anomalous data or distort it to match their personal beliefs (Zeidler, 1997). As a result, students may be able to construct elaborate arguments consisting of several relevant justifications that include accurate scientific knowledge, but their claim can be inaccurate because they did not attempt to coordinate their claim with all available evidence. This is a significant constraint to this method, especially because justifications for scientific claims are often based on interpretations of data gathered across multiple experiments.

Kelly and Takao's framework: Examining the epistemic status of propositions

Kelly and Takao (2002; Takao & Kelly, 2003) developed a method to analyze longer and more complex written arguments by examining term papers produced by students enrolled in an oceanography course. The term paper required students to support an abstract theoretical conclusion based on multiple data representations. The arguments generated by these students often contained multiple propositions in order 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 by the author to form a persuasive argument. In order to develop this framework, Kelly and Takao relied heavily on rhetorical studies of science writing (e.g., Bazerman, 1988; Latour, 1987). To analyze an extended rhetorical argument using this framework, propositions are identified and then sorted 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 the particular domain. Once classified, Kelly and Takao determine how these propositions are linked together and use this information to produce a graphical representation of an argument that shows how students coordinate propositions in their writing.

One limitation of the framework involves the lack of appraisal of how sensible the links between propositions are or even if the propositions are scientifically accurate. Without this type of evaluation, it is difficult to determine if the author understands the theories they apply or how well the data supports the conclusion. 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 and the appropriateness of how they were linked or it could be the result of a lack of sufficient evidentiary support for the students' conclusions (which are all field-dependant features an argument). Kelly and Takao's work once again illustrates the importance of evaluating the quality of arguments based on the standards developed within the discipline rather than solely relying on field-invariant structures (Toulmin, 1958).

Despite these limitations, Kelly and Takao's framework shows promise as an analytical tool. The examination of the epistemic status of the knowledge claims provides a way to characterize the types of propositions students use to support their conclusions and the extent to which individuals adhere to genre conventions of the scientific community. The development of this framework also highlights the importance of considering the nature of the task and the rhetorical aspects of argument construction when choosing or developing an analytical method to assess argument quality. For example, Kelly et al. (1998) found that students tend to rely on a series of relatively short arguments consisting of only a single claim as they attempt to persuade a lab partner of the accuracy of their ideas. In this type of task, Toulmin's framework works relatively well because arguments tend to be simple and the rhetorical aspects of the discourse are straightforward. On the other hand, the nature of the arguments found in the oceanography midterm papers examined by Kelly and Takao were much more complex. In these papers, students needed to formulate a line of evidence based on multiple data representations, which required students to reference specific data, identify specific geological features, and then explain how these features related to their abstract theoretical conclusions. Using Toulmin's or Zohar and Nemet's framework to analyze these arguments would have neglected important aspects of the rhetorical nature of these complex scientific arguments, such as the importance of moving rhetorically 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 indicates that the nature of the academic task will influence the type of argument developed by an individual and consequently the analytical methods needed to assess it.

Sandoval's framework: Examining the conceptual and epistemic quality of arguments

Sandoval (2003; Sandoval & Millwood, 2005) has developed an alternative framework for judging the quality of scientific arguments generated by students. Rather than examining arguments based on the field-invariant structural components of arguments, his coding scheme attempts to assess how well students generate arguments based on field-dependent criteria. Specifically, Sandoval's coding scheme assesses two dimensions of scientific arguments. First, conceptual quality measures how well the individual has (a) articulated causal claims within a specific theoretical framework, and (b) warranted these claims using available data. Second, epistemological quality measures how well the individual has (a) cited sufficient data in warranting a claim, (b) written a coherent causal explanation for a given phenomenon (see Sandoval, 2003) and (c) incorporated appropriate rhetorical references when referring to data (see Sandoval & Millwood, 2005). This analytical framework is particularly useful for two reasons. First, it determines if students can generate an argument that explains a particular observed phenomenon using a specific theory, such as natural selection. Second, it provides information about the epistemological criteria students use when generating arguments as an end product of their own inquiry and how these criteria align with the criteria used within particular scientific domains.

Sandoval's coding scheme suggests that constructing high-quality arguments requires (a) a conceptual understanding of relevant scientific theories and their application to a specific problem as well as (b) an epistemic understanding of the criteria for high-quality arguments. Sandoval argues for the importance of the latter component because the manner in which students incorporate and refer to data in their writing reflects their implicit epistemological commitments about the nature and role of data in the generation and evaluation of scientific knowledge. For example, in Sandoval and Millwood's (2005) analysis indicates that students are able to apply their understanding of natural selection to generate an argument that is consistent with the major tenets of natural selection. However, the overall pattern of warrant and evidence citation suggests that although students understand the importance of linking evidence and claims, students tend to rely on a single piece of data when supporting a particular claim. As a result, students often do not include a comparison of data from multiple sources when warranting a claim even when such comparisons are needed. Sandoval's (2003) work also indicates that students often interpret data incorrectly even though they can articulate a specific explanation in terms of a guiding theory. Taken together, this analytical framework highlights the importance of examining scientific arguments in terms of field-dependent criteria.

Lawson's framework: Examining the hypothetico-deductive validity of arguments

Lawson (2003) argues that science educators should 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 "determine which of two or more proposed alternative explanations (claims) for a puzzling observation is correct and which of the alternatives are incorrect" (p. 1389, emphasis in original). This process requires the generation of an argument that consists of not only a tentative explanation that may be correct but also includes how this explanation was tested 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 another.

The process of constructing a hypothetico-predictive argument begins with an observation that provokes a casual question and the generation of one or more tentative explanations. Once generated, these explanations must be tested in order 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" are used to 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. This match or mismatch of evidence and prediction can then be used to draw a conclusion regarding the validity of the explanation. Lawson indicates that the overall quality of this type of argument should be evaluated based on its 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.

Synthesis, conclusions, and recommendations for improving the analytical methods used to study rhetorical arguments in science education

As previously mentioned, scientific arguments are grounded in certain logical and epistemological commitments that are valued by the scientific community. Epistemological commitments refer to the philosophical assumptions and beliefs about the nature and limits of knowledge and its acquisition (Kitchener, 1983). These shared commitments make science different from other ways of knowing (Hogan & Maglienti, 2001; Sandoval, 2003). Consequently, several science educators have argued that students need to learn how the scientific community uses arguments in order to construct knowledge and the criteria for what counts as a good argument in science (e.g., Driver, Asoko, Leach, Mortimer, & Scott, 1994; e.g., Sandoval & Reiser, 2004). Unfortunately, as we have outlined in this article, the majority of the analytical methods that have been developed to assess and characterize the nature of the rhetorical arguments generated by students in science classrooms have provided very little information about how the rhetorical arguments generated by students reflect these criteria.

Although there is no single answer to the question of what constitutes a good argument in science, an operational definition for quality scientific argumentation would facilitate research and curricular development. As Osborne and colleagues (2003) suggest, using a generic framework to characterize argumentation across different science content areas and contexts can help students achieve a *basic understanding* of the processes and practices of science because there are several criteria that transcend specific disciplines and make science unique from other ways of knowing. For example, all scientists rely on evidence in order to justify ideas or knowledge claims because sooner or later the validity of scientific claims is settled by referring to observations of phenomena (Driver et. al., 2000). As Driver and colleagues (1994) suggest, scientific argumentation "is characterized by the explicit formulation of theories than can then be communicated and inspected in light of evidence" (p. 8). Once individuals develop this basic understanding, individuals can then learn the discipline specific practices and norms for creating and supporting knowledge claims within a particular domain. Nevertheless, developing a consensual view of quality argumentation from a scientific perspective requires linking research and theoretical perspectives on argument from various fields including, educational research, socio-cultural studies, literacy studies, history of science, and philosophy of science. Calls for such a link have been made before, yet little has been done. We therefore outline five criteria for examining the quality of scientific arguments that we view as essential *as a first step towards this goal*.

Criteria 1: Examine the nature and quality of the knowledge claim

In scientific arguments, claims are explanatory conclusions or descriptive frameworks rather than definitions or personal opinions. Scientists do not simply form context-specific conclusions or describe their observed results; they work to develop generalizable explanations for why something happens or descriptions of how something happens. Unfortunately, few students see science as a process of creating and testing frameworks and theories; instead science is seen as a steady accumulation of facts about the world. For example, Driver et al. (1996) found that students tend to describe scientific claims as "observations about the world" instead of descriptive frameworks and explanations that can be tested empirically. Similarly, Carey et al. (1989) found that students do not distinguish between experimental findings and the ideas that experiments are designed to test. Research also suggests that students rarely provide causal mechanisms for an observation based on an interpretation of data unless they are explicitly told to do so or they are scaffolded in a way that requires them to do so (Sandoval & Reiser, 2004). Taken together, this research suggests that students must learn more about the kinds of claims scientists make and how they advance them. Analytic methods should therefore focus on the type of claims students make as a result of their inquiry and whether or not the claims answer the research questions that guide their inquiries. Finally, assessments should also make judgments concerning the conceptual quality of these claims. The accuracy of a student's claim from a scientific perspective can revel a great deal about how a student understands content and how well the student is able to coordinate claims with available evidence.

Criteria 2: Examine how (or if) the claim is justified

Several studies have shown that although students are able to generate claims to explain data when asked to do so, students usually do not justify their claims unless they are challenged (Jimenez-Aleixandre et al., 2000; Kelly et al., 1998). When warrants are included as part of an argument, many people tend to rely too heavily on unsubstantiated explanations to justify their claims (Kuhn, 1993) or simply use plausible explanations as a way to replace evidence that is lacking or missing (Brem & Rips, 2000). In science, empirical evidence is an important consideration for evaluating the validity of scientific claims (Jimenez-Aleizandre, Rodriguez, & Duschl, 2000), therefore students must learn to not only provide empirical evidence in support of their ideas but also learn what kinds of evidence is needed to warrant an argument and how that evidence can be appropriately interpreted given community standards. Analytic methods should not only focus on whether or not students are justifying their claims but also on how individuals attempt to accomplish this task. In addition, assessments should also make judgments concerning the types of reasons and evidence that are used to support a particular claim. Do students rely on an appeal to authority figure in order to justify their ideas? Do they understand the difference between correlation and causation? Do students rely on empirical data or personal experiences as evidence? The types of reasons and evidence used by students and how they use them can tell us a great deal about what they consider good explanations or warranted knowledge.

Criteria 3: Examine if a claim accounts for all available evidence

A central activity of a scientist is to construct arguments determining which of the various conjectures about a puzzling phenomenon are most convincing in light of all available evidence and to obtain additional evidence when the available evidence is insufficient or lacking (Lawson, 2003). Unfortunately, a great deal of research suggests that students often fail to attend to important patterns in data and tend to preferentially consider a single piece of evidence rather than a set (Driver et al., 1996). Research also suggests that students often select, ignore, or distort evidence (Zeidler, 1997) especially when it threatens strongly held ideas. As a result, students are often able to generate arguments that consist of an inaccurate claim that is supported by empirical evidence because students tend to rely on a single piece of evidence while ignoring other data (Clark & Sampson, 2005). These findings indicate that analytical methods should also include a judgment of how well an individual is able to coordinate their claim with all available evidence.

Criteria 4: Examine how (or if) the argument attempts to discount alternatives

Students need to understand that scientific theories and conclusions are human constructs. As it often happens in science, more than one claim may be an acceptable explanation for the same phenomenon. Arguments are the preferential tool for resolving controversies in science, but when scientists have to make choices, evidence is never totally determinate, nor are arguments overwhelmingly convincing. More than one alternative is not just possible but often also plausible. As a result, students should learn to challenge alternative explanations for the same phenomenon by addressing weaknesses in other possible explanations or frameworks as a part of their argument. Unfortunately, research suggests that students are poor at presenting arguments ‘for and against’ or examining different points of view on the same issue (Driver et al., 2000). Analytical methods should therefore evaluate an individual’s ability to “postulate possible interpretations and then examine the arguments for each in the light of the evidence that is available to them” (Driver et al., 2000, p. 299). Although it is not always possible to exclude all but one explanation in science, students should be able to assess the acceptability or usefulness of alternative explanations using criteria such as: (a) the ability of the claim to account for all observations relating to the phenomenon, (b) its usefulness in predicting the behavior of the phenomenon, and (c) the consistency of the claim with the other accepted knowledge about how the world works (Driver et al., 2000; Passmore & Stewart, 2002).

Criteria 5: Examine how epistemological references are used to coordinate claims and evidence

As students engage in the construction of arguments they must do more than simply justify their ideas with evidence, they must also learn to justify that the evidence was appropriately gathered and interpreted given community standards. Therefore when constructing an argument that can be used to persuade, the ability to justify a claim by including a discussion concerning, for example, the validity of an experimental design or the interpretation of data is extremely valuable. Unfortunately, current research suggests that students’ understanding of the epistemological commitments of science, especially an understanding of what counts as warranted scientific knowledge and what methods can satisfactorily generate such knowledge is poor (Sandoval & Millwood, 2005). Students frequently indicate that the results of an experiment are enough to prove ideas right or wrong without ever questioning the methods used in the experiment or how the data was analyzed (Carey et al., 1989). Working to help students address the validity and reliability of their evidence, as well as the presence of other confirming or disconfirming evidence, will be important if students are to generate productive scientific arguments.

In conclusion, a key question we should ask about any analytic framework of rhetorical argumentation concerns what the framework can tell us about the quality of the arguments generated by students when they are asked to explain or describe a complex phenomenon. Structural analyses of arguments have made explicit the difficulties students encounter in marshaling evidence, drawing on their conceptual understanding of the topic, and composing arguments in support of a scientific knowledge claim. The epistemic demands of argument construction, however, engage the criteria to which knowledge claims are held within a community of practice. Although it is important to help students learn to construct elaborate arguments in order to facilitate the development of their thinking and informal reasoning skills across contexts (Osborne et al., 2004), we must also recognize the significance of enabling learners to engage in argumentation in a way that reflects the standards used to construct and evaluate arguments by the scientific community. Ultimately, this analysis strongly suggests the need for further development and improvement of the tools used to study argument in science education. The five criteria for the assessment of rhetorical arguments outlined in this article are a first step towards this goal.

References

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

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

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

Clark, D. B., & Sampson, V. (2005). *Analyzing the quality of argumentation supported by personally-seeded discussions.* Paper presented at the Computer Supported Collaborative Learning Conference, Taipei, Taiwan.

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

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Philadelphia: Open University Press.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-313.

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

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. *Human Development, 26*, 222-232.

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

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: Harvard University Press.

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.

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.

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.

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*.

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

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: Cambridge University Press.

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

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.