Chapter 3
Methodological Foundations in the Study of Argumentation in Science Classrooms

Sibel Erduran

> *"Every discourse, even a poetic or oracular sentence, carries with it a system of rules for producing analogous things and thus an outline of methodology."*

<div align="right">Jacques Derrida</div>

Ask anyone who has done work on argumentation in science classrooms what their primary concern has been in this line of research, and they will most likely respond with one word: *methodology*. Most likely they will then begin to ask you if you have figured out how to distinguish *data* from *warrants*. The questions will continue: can theoretical statements be data? If a warrant is not explicitly stated, can it still be assumed that it is part of the argument? Indeed the study of argumentation in the science classroom raises significant methodological questions. What counts as an argument in children's talk anyhow? What is the unit of analysis of argument and of argumentation in classroom conversations? What criteria drive the selection and application of coding tools? What justifies the choice of one methodological approach over another? What does a particular methodological approach enable us to do and how does it do so?

While in one sense, such methodological questions are about the reliability and validity of methodological tools for the analysis of arguments (e.g., Duschl et al., 1999), in another sense they are questions about the very nature and function of methodologies for a line of research that challenges positivist characterizations of scientific knowledge stripped off of the cultural, affective, economical and personal contexts and processes of science. In a review of literature on the use of methodologies in science education, Kelly et al. (1998) observed incongruities between theoretical perspectives and methodological approaches adapted in studies on the Nature of Science. Although the bodies of literature informing the Nature of Science studies used multiple methodological orientations, the majority of the empirical Nature of Science studies used either survey instruments or interviews, without observational data of teachers and students. The state of affairs in the case of argumentation might present an example of an opposite trend where, roughly two decades later since argumentation has taken root in science education, our methodological work remains heavily focused on observational data at the expense of surveys and interviews. It is worthwhile to note that concentrating on quantitative analyses of argumentation does not necessarily imply a contradiction between methodological and theoretical orientations of science education. Quantitative analyses

address different questions from those raised by detailed analyses of classroom talk. For instance, "what correlations are there between power relations in classrooms and the ability to argue scientifically?" is a question that begs a methodological orientation based on quantitative methods whilst at the same time empowering the sociological processes of science in the classroom.

This chapter will trace issues related to such methodological questions surrounding the study of argumentation in science classrooms particularly in an effort to provide a rationale for what methodological approaches enable science education researchers to accomplish and how. For example, methodological tools need to be refined enough to generate a set of indicators for the quality of arguments generated in the learning environment (Erduran et al., 2004). A further emphasis of the chapter will be on the application of particular theoretical frameworks (e.g., Toulmin, 1958; Walton, 1999) as well as the generation of categories from data-driven approaches (Sandoval & Reiser, 2004; Maloney & Simon, 2006). In so doing, the chapter will problematize the adaptation of theoretically and empirically grounded perspectives as methodological approaches, and it will investigate some challenges that such approaches can pose. Finally, the role of methodological innovations in contributing to the knowledge base in science education will be explored. In particular, the case of Stephen Toulmin's (1958) work will be used to illustrate what contribution the adaptation of his framework on argument has made to knowledge in science education. A significant body of argumentation literature in science education has been based on Toulmin's work (e.g., Erduran et al. 2004; Jiménez-Aleixandre et al., 2000).

It is interesting to note that even though science education as a field remains minimally influenced by philosophical analyses (Scerri, 2002), the uptake and impact of Toulmin's framework on argument (particularly as a methodological tool) has mirrored trends within philosophy itself. In "A Citation-Based Reflection on Toulmin and Argument", Ronald P. Loui (2005) uses citation counts to measure the influence of Toulmin's work. He reports that citations in the leading journals in the social sciences, humanities and science and technology put Toulmin and his works in the top 10 among philosophers of science and philosophical logicians of the 20th century. Thus, he concludes, Toulmin's *Uses of Argument*, and work in general, have been essential contributions to 20th-century thought. Even though there has been no quantitative measures of the impact of Toulmin's work in science education, qualitatively it would be difficult to disagree with the position that Toulmin's work has influenced the work of many science educators has had in the literature. Prevalence of Toulmin's work in application to the study of argumentation in science classrooms (e.g., Erduran et al., 2004) will be used as a case example of how methodological approaches can contribute to the development of knowledge in science education.

Analysis of Argumentation in School Science

In the 2003 Conference of the European Science Education Research Association, I was asked to be a discussant for a session titled "Communication and Discourse Analysis in the Science Classroom." The session included five papers and used a range of theoretically driven analytical frameworks for the study of discourse in science

classrooms. The work from a couple of these presentations has subsequently been published. Jiménez-Aleixandre and Pereiro Muñoz (2005) used Toulmin's framework to study students' interactions in small groups. Castells et al. (2007) used Perelman's Theory of Argumentation to frame teacher–student interactions from both epistemological and communicative perspectives. Marquez, Izquierdo and Espinet (2006) used Halliday's model of Functional Grammar to interpret communicative and linguistic aspects of teachers' actions. Piccinini and Martins (2005) used Kress and colleagues' semiotic modes to interpret teacher–student interactions. Scott and Mortimer (2005) drew on sociocultural perspectives including the work of Lev Vygotsky to study a range of interactions in the classroom including student–student interactions. An overarching theme across these papers was the assumption that there are teaching and learning situations that can be captured in semiotic interactions and that the study of semiotic interactions can inform and improve science education.

This conference session embodies some of the methodological issues in the study of argumentation particularly the adaptation of a certain theoretical stance from a leading scholar in a related field such as philosophy and linguistics. In a similar spirit as the ESERA session, literature on argumentation in science education has witnessed the adaptation of theoretical perspectives for methodological use (Erduran et al., 2004; Zohar & Nemet, 2002; Jiménez-Aleixandre et al., 2000) as well as the generation of analytical tools from a more *grounded* approach (Sandoval & Reiser, 2004; Maloney & Simon, 2006). The particular rationalization of these tools is done relative to the context of the research in which the tool was used and the purpose of the study. In the next few sections, I will review some of these approaches. In particular, I will illustrate how studies have focused on the analysis of (a) evidence and justifications; (b) epistemic practices and criteria; (c) arguers and the nature of arguments; and (d) participation in discussions, as criteria for defining and confining the analytical boundaries for argumentation in the science classroom.

Evidence and Justifications

Zohar and Nemet (2002) modified Toulmin's Argument Pattern (TAP) 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). 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, an approach also employed by Erduran et al. (2004). The criteria for the classification of justifications were (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. Zohar and Nemet's framework does not evaluate the accuracy of the claim itself. As a result, their framework works better when used to analyze arguments generated in the context of socio-scientific issues rather than in the context of scientific debates. In response to socio-scientific 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.

In our work (Erduran et al., 2004), we developed two methodological approaches for the analysis of discourse from whole class and small group discussions. First, we adapted TAP for the purposes of coding data that originate from whole-class conversations where successive implementation of lessons can be traced for their improved quality of argumentation. Here we have traced the frequency of TAP profiles from the same lessons that were implemented a year apart by the same teachers. Comparison of the results held the potential to investigate whether or not there was an improvement in the employment of argumentation across different lessons. Our purpose was not to report on statistically significant outcomes since our sample size was small (i.e., two lessons per teacher and no control lessons) but rather our aim was to describe a methodology that can be of use to future researchers in the quantification of arguments to test the effectiveness of interventions based on argumentation.

Our analysis provided a qualitative indication also of how teachers' specific discourse practices compare and thus how appropriate feedback can be crafted to facilitate particular teachers' implementation of argumentation. For example, the distribution of TAP profiles across the two years was very similar for each teacher but different between teachers. The tool we have developed, then, provided us with an insight into how teachers' engagement in argumentation compares and where in discourse more emphasis is needed to improve the quality of argumentation. We were also able to trace cross- and within-teacher variations in how argumentation was implemented (Simon et al., 2006). Given the research evidence that teachers' practices improve when they are empowered by reflection and understanding on their teaching actions (e.g., Loucks-Horsley et al., 1998) such insight could help create powerful strategies for more effective implementation of traditionally unfamiliar discourse forms such as argumentation.

A further outcome of our methodological approaches was a scheme reproduced in Table 3.1 where argumentation is assessed in terms of levels of the quality of oppositions or rebuttals in the student discussions in small-group format (Erduran et al., 2004). In this approach, we have focused on those instances where there was a clear opposition between students and assessed the nature of this opposition in terms of the strength of the rebuttals offered. We perceived the presence of a rebuttal as a significant indicator of quality of argumentation since a rebuttal, and how it counters another's argument forces both participants to evaluate the validity and strength of that argument. Research evidence (e.g., Kuhn, 1991) suggests that the cognitive skill of argument is, to some extent, founded on an understanding of how

Table 3.1 Analytical framework used for assessing the quality of argumentation (Erduran et al., 2004)

Level 1:	Level 1 argumentation consists of arguments that are a simple claim versus a counterclaim or a claim versus a claim.
Level 2:	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants or backings but do not contain any rebuttals.
Level 3:	Level 3 argumentation has arguments with a series of claims or counterclaims with either data, warrants or backings with the occasional weak rebuttal.
Level 4:	Level 4 argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counterclaims.
Level 5:	Level 5 argumentation displays an extended argument with more than one rebuttal.

to rebut an opposer's point of view. In this sense, students' ability to formulate strong rebuttals is a significant goal for the teaching of argumentation.

We thus traced the quality of argument by focusing on the presence or absence of rebuttals. For instance, when there was opposition between students but the opposition consisted of only counterarguments that were unrelated, we perceived this to be low-level argumentation. In other words, in these cases, there was no indication of an understanding of a rebuttal in terms of its relation to challenging the validity of the evidence and justifications offered. There was simply no reference to the components of the argument maintained by the opposition. When, however, the rebuttal was in direct reference to a piece of evidence (data, warrants or backings) offered, thereby engaging with a presented argument, we considered this instance to be representative of higher level argumentation. In this methodological approach, we have thus emphasized the use of rebuttals and developed a strategy for using TAP as a measure of interactive discourse.

Epistemic Practices and Criteria

Kelly and Takao (2002), and Takao and Kelly (2003) developed a method to analyze longer and 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 focused on the relative epistemic status of these propositions and how these propositions were 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.

Sandoval and Millwood (2005) have developed a framework for judging the quality of scientific arguments generated by students. Rather than examining arguments based on the field-invariant structural components of arguments, these authors' coding scheme attempts to assess how well students generate arguments based on field-dependent criteria. Specifically, Sandoval and Millwood'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, and (c) incorporated appropriate rhetorical references when referring to data.

A strength of Sandoval and Millwood's framework is that it can determine if students can generate an argument that explains a particular observed phenomenon using a specific theory, such as natural selection. Furthermore, their framework 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 and Millwood's scheme suggests that constructing high-quality arguments requires a conceptual understanding of relevant scientific theories and their application to a specific problem as well as an epistemic understanding of the criteria for high-quality arguments. These authors argue 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, 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 where when such comparisons are needed. Sandoval's earlier work (Sandoval, 2003) also indicates that students often interpret data incorrectly even though they can articulate a specific explanation in terms of a guiding theory. A key contribution of Sandoval and colleagues' analytical frameworks is the observation that field-dependent criteria are important in the analysis of arguments.

Arguers and Nature of Arguments

Zembal-Saul et al. (2003) developed a rubric to analyze pre-service teachers' arguments (Table 3.2). The rubric consisted of four main categories: causal coherence and structure; evidence; justifications and evaluations. In a qualitative case study,

Table 3.2 Rubric for analyzing pre-service teachers' arguments (Zembal-Saul et al., 2003)

1. Causal Coherence/Causal Structure
   (a) A network representation of causal relations was constructed based on students' explanations.
   (b) Description of the causal sequence
       (i)   Do explanations articulate specific cause-and-effect relationships?
       (ii) Are causal relationships logically connected?
       (iii) Are causal relationships and their connections explicitly stated?
       (iv) Do they consider the possibility of more than one cause (multiple causal lines)?
   (c) Do they consider the possibility of multiple factors interacting to produce a phenomenon?
   (d) Does the causal structure reflect domain-specific principles (e.g., selective pressure, change in frequency traits in population, initial variation, differential survival)?
2. Evidence
   (a) Is there evidence to support each claim?
   (b) Is the evidence relevant to the claim?
   (c) Do they make valid inferences from data?
   (d) Do they use principles of knowledge within the domain?
   (e) Do they sort data in appropriate ways (e.g., based on population characteristics such as sex and age)?
   (f) In which cases do they have more or less pieces of data linked as supporting evidence? What distinguishes parts that are supported with several pieces of evidence and those that are not?
   (g) Do they tend to use individual data or representations of population patterns such as graphics? In what circumstances do they use different kinds of evidence?
   (h) Do they tend to use qualitative data or quantitative data to support their claims? In what circumstances do they use different kinds of evidence?
   (i) How do they describe their pieces of evidence (e.g., annotation box in software)? Do such descriptions vary depending on the type of evidence (e.g., graphs, field notes)?
   (j) Is it possible to identify any changes in these aspects across the unit (e.g., when do they start to use a type of evidence?)?
3. Data justifications
   (a) Do students provide justification for why data is relevant to support a claim?
   (b) What kind of justification do they use?
   (c) Are there particular instances in which justification is absent/present?
4. Thinking about their explanations (evaluating their explanations)
   (a) How do they categorize their explanations (e.g., accepted completely; accepted with changes)?
   (b) How do they justify this categorization?

pre-service science teachers enrolled in their advanced methods course participated in a complex, data-rich investigation. Fundamental to the investigation was the use of the *Galapagos Finches* software and an emphasis on giving priority to evidence and constructing evidence-based arguments. The primary sources of data were the electronic artifacts generated in the *Galapagos Finches* software environment and the videotaped interactions of both pairs as they investigated the data set, constructed and revised their arguments, engaged in peer review sessions, and presented

their arguments to the class at the end of the unit. One of the outcomes of the case study was that using the software pre-service science teachers consistently constructed claims that were linked to evidence from the investigation. Another outcome was that although pre-service science teachers consistently grounded their arguments in evidence, they still exhibited a number of limitations reported in the literature.

Hogan and Maglienti (2001) developed a coding scheme for rating participants' overall judgment of a conclusion (Table 3.3). These researchers examined the criteria that middle school students, non-scientist adults, technicians, and scientists used to rate the validity of conclusions drawn by hypothetical students from a set of evidence. The groups' criteria for evaluating conclusions were considered to be dimensions of their epistemological frameworks regarding how knowledge claims are justified, and as such how they are integral to their scientific reasoning. Quantitative and qualitative analyses revealed that the responses of students and non-scientists differed from the responses of technicians and scientists, with the major difference being the groups' relative emphasis on criteria of empirical consistency or plausibility of the conclusions.

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). 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

Table 3.3 Coding of conclusions (Hogan & Maglienti, 2001)

Level	Description
0	Does not mention any relevant strengths and weaknesses of the conclusion.
1	Mentions some relevant strengths and weaknesses of the conclusion, but not the major ones. Also uses agreement with personal inferences or views as a basis for judging the conclusion.
2	Mentions some strengths and weaknesses of the conclusion, but not the major ones. Does not base judgments on agreement with personal inferences or views.
3	Mentions the major strengths and weaknesses of the conclusion, but also uses agreement with personal inferences or views as a basis for judging the conclusion.
4	Mentions the major strengths and weaknesses of the conclusion. Does not base judgments on agreement with personal inferences or views.

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.

Participation in Discussions

Maloney and Simon (2006) developed a coding system to show different approaches to engaging in discussion. The system, termed a "Discussion map", was designed to identify the nature and extent to which children engaged in sustained argumentation dialogue. The construction of these maps was initially informed by the work of Chinn and Anderson (1998), who used "argument networks" to analyze the structure of discourse of children in small groups as they discussed issues raised by stories (not scientific in nature). One of the major problems encountered in the use of argument networks was of a practical nature; a transcript that was 4 pages in length produced an argument network that required 13 pages. However, the construction of argument networks identified the need for some diagrammatic representation of the discourse, as the diagrams demonstrated clearly the varying patterns of discussion for the different activities. For example, they showed whether the arguments put forward were discussed by the group or ignored, and whether arguments were followed by the presentation of a new claim. For opposing arguments, the diagrams indicated whether the evidence was examined to evaluate the opposing claims or whether claims were just accepted and not challenged.

The diagrams also showed which children were taking part in the discussions. As a result of this pilot work and the developing clarity about the requirements to aid analysis, the "Discussion Map" was devised to capture all these features in a more economic way. A Discussion Map is constructed through identifying key episodes of "talk" that include argumentative discussion using evidence. These episodes are termed "Argument", "Review", and "Clarification". A fourth category of talk was needed to complete the transcript analysis, so that the Discussion Map captures the

intervals and frequency of the key episodes of talk—this fourth category includes all other types of discourse and is termed "Other Talk".

Trends in the Literature on Analysis of Argumentation

The preceding review of literature is not intended to be exhaustive and thus cannot be used as a definitive source for analytical perspectives on argument in science education. However, it is noteworthy to state that the pattern in the use of analytical frameworks to study argumentation in school science has tended to emphasize the qualitative aspects of the structure of an argument and the processes of argumentation. Given the labor-intensive nature of analysis of classroom and group talk, this observation is not surprising. In our work (Erduran et al., 2004) we attempted to develop quantitative measures of the quality of argumentation and yet it is unlikely that our methodological approach can be realistically adapted for large-scale studies. Neither should they be if the questions that such methodology targets make large-scale quantitative measures meaningless. Consider the task of a biochemist who is interested in the particular features of a protein, perhaps how certain amino acid sequences might dictate the function as an enzyme. It would be meaningless to generalize or to quantify the features of such sequences to all enzymes given their particular functions. In other words, it is the particular nature of the object of study that is of interest and that guides the research question.

Whilst it is important to focus on discourse to illustrate the nature of argumentation and reasoning (for further reference on a review of analytical approaches, see Clark et al., in press), it is equally important to introduce methodological approaches that aim at addressing different questions, particularly questions that seek understanding of correlations and associations. For instance, a question such as "is there a significant impact of argument skills on subject knowledge in science?" would necessitate that tools are generated and applied to data to measure both argument skills and subject knowledge, and that the joint use of these tools can be justified.

Furthermore a significant deficit in the literature remains which is the paucity of research on quantitative analysis of argumentation, not at the level of conversational analysis but at the level of conceptual categories that are of significance to science education. For example, there is limited understanding of how teachers' beliefs about pedagogical values of discussions might correlate with their emphasis in their teaching of argumentation. Likewise there are no measures of teachers' and students' attitudes and beliefs about the role of argument in science and in science education. One exception to this overall pattern is the work of Sampson and Clark (2006) who have developed a questionnaire to assess the correlation between argumentation skills and understanding of the nature of science. Overall, however, the trends in the literature point to the challenges that researchers have experienced in the qualitative analyses of argumentation in the science classroom which is the focus of our discussion in the next section.

Coding Arguments: Challenges and Compromises

A major issue in the study of argumentation in either written or verbal data is the unit of analysis. What becomes of the boundary markers of the data where arguments begin and end? Decisions have to be made regarding how the data will be split and subsequently how the chunks will be categorized and interpreted. Is an argument located within one person's argument or would a set of statements still count as the components of an argument even if the talkers may not have intended them to be part of a bigger whole? Let us explore such questions taking on a definition of argument based on Toulmin's work.

Toulmin's Argument Pattern (TAP) (Fig. 3.1) illustrates the structure of an argument in terms of an interconnected set of a claim; data that support that claim; warrants that provide a link between the data and the claim; backings that strengthen the warrants; and finally, rebuttals which point to the circumstances under which the claim would not hold true. More specifically, a claim is an assertion put forward publicly for general acceptance. Data and warrants are the specific facts relied on to support a given claim. Backings are generalizations making explicit the body of experience relied on to establish the trustworthiness of the ways of arguing applied in any particular case. Rebuttals are the extraordinary or exceptional circumstances that might undermine the force of the supporting arguments. Toulmin further considers the role of qualifiers as phrases that show what kind of degree of reliance is to be placed on the conclusions, given the arguments available to support them. (Toulmin's framework will be discussed in more detail in a later section of this chapter.)

Despite its use as a framework for defining argument, the application of TAP to the analysis of classroom-based verbal data has yielded difficulties. The main difficulty has been in the clarification of what counts as claim, data, warrant and backing. Kelly et al. (1998) applied TAP to the analysis of student dyadic spoken discourse. This study identified the potential uses of Toulmin's method but also
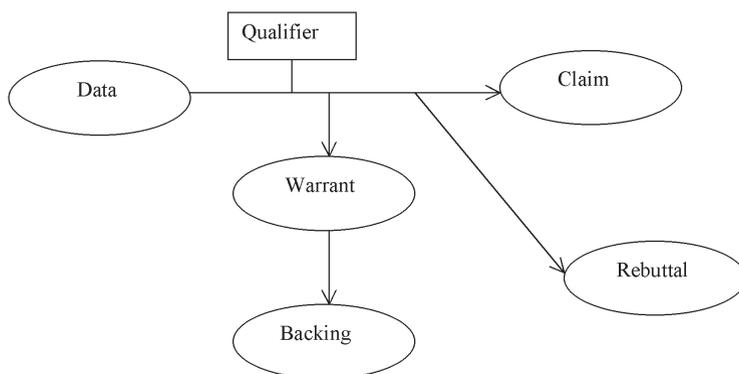

Fig. 3.1 Toulmin's argument pattern (Toulmin, 1958)

highlighted some methodological problems. The authors found that organizing student discourse into Toulmin's argument components required careful attention to the contextualized use of language. According to Kelly and his colleagues, while the Toulmin model makes distinctions among statements of data, claim, warrant and backing, the scheme is restricted to relatively short argument structures and the argument components pose ambiguities. Statements of claims can serve as a new assertion to be proven or can be in service to another claim, thus acting as a warrant.

In a subsequent study, Kelly and Chen (1999) modified Toulmin's model by drawing on the work of Latour (1987). They thus considered the epistemic status of students' claims in their writings and sorted these according to the model presented by Latour. This form of analysis allowed for the consideration of claims at multiple levels of theoretical generality and matched well with the categorical description of transactional use of language. Other researchers (see Duschl this book) have preferred to use other analytical tools such as Douglas Walton's scheme on presumptive reasoning, justifying their choice on the ambiguity surrounding the key features of TAP in application to real discourse.

Let's illustrate the difficulties encountered in the coding of arguments using TAP with an example. The episode comes from our published work (Erduran et al., 2004) conducted in a middle school in London where we explored children's argumentation in whole class discussions. The students were asked to evaluate a set of statements regarding the phases of the moon.

Teacher	Statement A, "The moon spins around, so the part of the moon that gives out light is not always facing us." Julian, A?
Student	The moon doesn't give out light.
Teacher	Right, so that's why A is wrong. That's true. How do you know that?
Student	Because the light that comes from the moon is actually from the sun.
Teacher	He is saying the light that we see from the moon is actually a reflection from the sun. How do we know that?
Student	Because the moon is blocked by the......

In our earlier work, we have used this example to illustrate how we resolved some of the issues involved in coding. I will reconstruct this example here to illustrate some of the key challenges that we faced in coding arguments in our work. In this example, one could consider the statement "The moon spins around" as a piece of data that supports the claim "So the part of the moon that gives out light is not always facing us." One could also argue, however, that the student's choice of "A" (the statement on the card) is the main claim. In other words, "A is right" can be considered an implicit claim that is challenged by the next claim "The moon doesn't give out light." Deciding which of statements to take as a claim (i.e., "The moon spins around" or "A is right") can thus become problematic.

Examining the use of words such as "so" and "because" can help resolve some ambiguities. Indeed, the use of the operative word "so" which itself is implied in Toulmin's definition (for reaching conclusions from data) makes the first case described highly convincing. In other words, there is little doubt that there is a claim and a justification, whatever the precise nature of this justification might be

or indeed whichever statement ("The moon spins around" or "The part of the moon that gives out light is not always facing us") is taken to be the main claim. The use of the next statement "The moon doesn't give out light" as a rebuttal creates an opposition to the justification used in the primary argument. The student's further elaboration of reasoning in "Because the light that comes from the moon is actually from the sun" is an effort for a justification of the rebuttal. Viewed in this way, ambiguities about what counts as claim, data, rebuttal and so on can be resolved. Even though all the statements above can be considered as claims in themselves, in the course of the reasoning, they can be positioned to be data or rebuttal relative to the main claim that creates an impetus for the generation of the subsequent statements. Indeed many aspects of an argument can be considered "nested" where, for instance, data of one argument could count as a claim for another argument.

Resolving such differences in coding is not a matter specific to analysis of argument. Establishing clearly defined and codable categories is a major issue in qualitative data analysis in general. As in any kind of analysis, a significant issue is that the categories have to be tight enough to be able to capture what we want them to capture. The nature of codes and the strategy for coding will depend on the purpose of the investigation as well as the questions that the research is trying to address. For the purposes of coding, the researcher would need to specify the instructions for new coders so as to ensure that reliable coding can occur. For example, in the preceding episode, a researcher may make a decision that the statement cards will be treated as main claims relative to which all other statements will be positioned.

Apart from ambiguities in what counts as a claim, data and warrant, other challenges exist for the coding of arguments. For instance, if the components of an argument are repeated, can we establish that a new argument is not introduced to count as another argument? Can we establish the role and function of such repetition in conversation? What if the student says a bit more in a sequence of talk? Would spatially separate but seemingly related statements count as parts of the same argument and add to the original argument? The researcher will also need to create boundaries and rules for such cases. A further issue is the nature of evidence used as data, warrants and backings. Can theoretical statements count as evidence or should evidence be empirically based? Can opinions, beliefs, ideas and values count as evidence? Is there a difference between what counts as evidence in scientific and socio-scientific contexts? The source of the components of the argument—i.e., whether or not they are empirical and theoretical—presents another problem for coding arguments. Researchers might be interested in examining the validity of arguments relative to the use of evidence. In certain respects the use of empirical evidence might be more favorable than theoretical evidence. In others, theoretical statements might be the only source of evidence, for instance, the use of the atomic theory in arguments about why a chemical reaction takes place in a particular way and not in another fashion.

A further challenge in the study of argumentation in science education is the extent to which codes of arguments can frame pedagogical and learning aspects of argumentation. Zembal-Saul et al.'s (2003) rubric for analyzing teachers' arguments rests heavily on their understanding and use of argument structure and process. In our work on the professional development of science teachers to include argumentation in their teaching, we developed a hierarchy of codes that are intended to capture the pedagogical

strategies underlying argumentation episodes (Simon et al., 2006). The subsequent investigation was to identify how the teacher might be promoting the implementation and learning of such concepts. For example, playing the role of devil's advocate could be considered as a pedagogical strategy that promotes the use of justifications. In both of these example cases, concepts such as evidence, claim and justification—central to the definition of argument—guided the focus on the text for analysis. In other words, an implicit entry point into the transcripts to examine teaching behaviors was a definition of argument. It is difficult to imagine what other entry point or a guiding framework could be used for this purpose other than a definition of argument. The precise intention is to seek to understand the nature of argumentation be it from a pedagogical, learning or any other point of view. What this observation does point to is the significance of which definition of argument is being used for pedagogical purposes and how such a choice on argument can dictate the analysis sought beyond just a definition of argument.

Revisiting Toulmin: Contributions of Methodology to Knowledge in Science Education

There is little doubt that Toulmin's seminal book *The Uses of Argument*, first published in 1958, has guided much research in science education. The preceding discussions provide evidence to this observation. I personally have been influenced by Toulmin's work for several years starting with the work we did in early 1990s in Pittsburgh schools on promoting scientific inquiry (Duschl, this book). Ever since, I made numerous attempts to contact Toulmin in the Department of Anthropology at the University of Southern California which have not yielded a response. I wanted to ask Toulmin himself what he thought about the way in which science educators have considered and adapted his work for educational purposes—a question that may be of interest to other researchers to pursue as well. Toulmin's model has been appropriated, adapted and extended by researchers not only in science education but also in the fields of speech communications, philosophy and artificial intelligence.

One issue of the journal *Argumentation* in 2005 brought together the best contemporary reflection in these fields on the Toulmin model and its current appropriation. The volume included 24 articles by 27 scholars from 10 countries. The papers extended or challenged Toulmin's ideas in ways that make fresh contributions to the theory of analyzing and evaluating arguments. Collectively, they represent the only comprehensive book-length study of the Toulmin model. They point the way to new developments in the theory of argument, including a typology of warrants, a comprehensive theory of defeaters, a rapprochement with formal logic, and a turn from propositions to speech acts as the constituents of argument.

As an illustration of his framework of argument, Toulmin (1958) discusses the claim that Harry is a British subject. The claim can be supported by the datum that Harry was born in Bermuda. That there is a connection at all between datum and claim is expressed by the warrant that a man born in Bermuda will generally be a

British subject. In turn the warrant can be supported by the backing (that there are certain statutes and other legal provisions to that effect). The warrant does not have total justifying force, so the claim that Harry is a British subject must be qualified: it follows presumably. Moreover there are possible rebuttals, for instance when both his parents were aliens, or he has become a naturalized American.

Verheij (2005) argues that since the appearance of Toulmin's book, the following ideas have found increasing support in different research communities (under the direct influence of Toulmin or independently): (a) in argumentation, the warrants of arguments (in the sense of inference licenses) can be at issue and their backings can differ from domain to domain; (b) arguments can be subject to rebuttal in the sense that there can be conditions of exception; (c) arguments can have qualified conclusions. (d) other kinds of arguments than just those based on the standard logical quantifiers and connectives (for all x, for some x, not, and, or, etc.) need to be analyzed; (e) determining whether an argument is good or not involves substantive judgments and not only formal logic.

In the next sections, I will review some of the contributions of Toulmin's work to themes related to science education including expert–novice studies, problem-solving, scientific reasoning, and theoretical representations and frameworks. My purpose here is to provide some examples of how a theoretically informed definition of argument can yield methodological approaches, which in turn can contribute to knowledge in the field of science education.

Contributions to Expert–Novice Studies and Problem-Solving

Cognitive scientist James Voss (2005) regarded the application of the Toulmin model as a success in the sense of providing a tool that produced a reasonable structure to complex protocols thereby enabling the study of expert reasoning in different domains. Voss used Toulmin's framework in the analysis of verbal protocols obtained during the solving of ill-structured problems. The research Voss conducted involved experts on the Soviet Union indicating how they would improve the USSR's poor agricultural productivity. Each expert was asked to assume he or she was Head of the Soviet Ministry of Agriculture and was asked how agricultural productivity could be improved. Each person responded orally, thus providing a "think aloud" protocol, the account being tape-recorded. For the analysis the Toulmin model was extended in order to enable description of lines of argument found in protocols as long as 10 paragraphs.

Results included that (a) while the protocol was comprised of a large number of specific arguments, the analytical approach enabled the tracing of a solver's line of argument; (b) on occasion datum and backing were difficult to distinguish; (c) warrants essentially were not stated, although substantial backing was provided. However, as perhaps would be expected, the Toulmin model did not provide for delineation of components of the problem-solving process. A second analysis assuming a "higher level" problem-solving structure and a "lower level" argument

structure produced an integrated problem-solving—argumentation structure depicting how reasoning is used in relation to particular task goals. Finally, at a more general level, problem-solving was considered as a classical rhetorical structure.

According to Voss what was especially gratifying was the reasonably clear lines of argument that were obtainable. Furthermore, the analysis led to other questions such as the nature of the protocol differences found not only among people of different knowledge but also among different experts. With respect to the actual coding of the protocols, Voss and colleagues experienced some difficulty determining whether a given statement was datum or backing, especially when a signal word such as "because" did not occur. A second finding was that, within their scoring system, the warrant of an argument was almost never stated. Following Toulmin (1958) as well as Hample (1977) and Govier (1987), statement of the warrant would make the argument logically valid. However someone who is solving a problem may only be interested in providing support for the claim and thus is concerned with backing, not deductive purity.

Voss indicates that the main shortcoming of the Toulmin model relates to the goal of studying solving ill-structured problems. While providing a means to isolate lines of argument, the Toulmin model did not provide information concerning the problem-solving process. One could of course argue that the Toulmin analysis was not designed for this purpose. Voss and colleagues did conduct a second analysis in an effort to enhance our understanding of the problem-solving process (Voss et al., 1983). They assumed the existence of a "higher level" problem-solving structure that included the elements of the previously described information processing model.

The protocol data were analyzed in relation to this structure, and the argument structure then became "lower level" with respect to the problem-solving structure. Moreover, two sets of operators were used in the analysis, one in reference to the problem solving structure and the other in relation to the reasoning or argument structure. The operators for the former were state constraint, state sub-problem, state solution, interpret problem statement, evaluate, and summarize. For the latter the operators were state argument, state assertion, state fact, present specific case, state reason, state outcome, compare and/or contrast, elaborate and/or clarify, state conclusion, and state qualifier. This analysis was of particular interest because it showed how specific arguments were employed in argument sequences or "nests of arguments" which in turn were employed in relation to higher-order problem-solving goal structures.

The contents of the problem-solving structure constitute an argument. In particular, the solver is faced with a question or problem and the representation phase essentially involves an analysis aimed at providing a statement of the cause(s) of that problem, with problem history often being part of this analysis. A solution is then proposed (usually experts prefer an overall solution whereas novices tend to list specific sub-problem solutions) and the remainder of the solution phase consists of justification of that solution. Thus, the solution is the claim, the datum consists of the causal factors, and the solution development constitutes backing. In agreement with Toulmin, this solving process places emphasis upon

justification (e.g., Van Eemeren et al., 1996). In the spirit of expert–novice studies, future analytical approaches will be enriched with foci on the development of expertise in science subject knowledge (e.g., Erduran, in press) and pedagogical content knowledge (e.g., Erduran & Dagher, in press). In other words, methodological frameworks that help investigate how science domain-specificity and levels of pedagogical competence relate to professional development will further contribute to the literature on expert–novice studies.

Contributions to Scientific Reasoning

A genuine and radical deviation from standard logic is required by Toulmin's notion of rebuttals although Toulmin hardly elaborates on the nature of rebuttals, a key reason why science educators such as ourselves (Erduran et al., 2004), for instance, have deviated from a formal definition of rebuttal in making TAP applicable for data analysis where rebuttals were involved.

As Toulmin puts it, rebuttals involve conditions of exception for the argument (Toulmin, 1958; p. 101). Apparently, for Toulmin, rebuttals can have several functions. For instance, rebuttals can "indicate circumstances in which the general authority of the warrant would have to be set aside" (p.101), but can also be (and for Toulmin apparently equivalently) "exceptional circumstances which might be capable of defeating or rebutting the warranted conclusion" (p. 101). On p. 102, he also speaks about the applicability of a warrant in connection with rebuttals. In other words, Toulmin speaks of the defeat (or rebutting) of the conclusion, of the applicability of the warrant and of the authority of the warrant, in a rather loose manner, without further distinction. Toulmin is unclear about the relation of these seemingly different situations. Here the three will be distinguished, in a way that naturally fits the reconstruction of the other elements of Toulmin's scheme above, as follows. If we look at the warrant–data–claim part of Toulmin's scheme there are five statements that can be argued against (Verheij, 2005):

1. The data D
2. The claim C
3. The warrant W
4. The associated conditional "If D, then C" that expresses the bridge from datum to claim.
5. The associated conditional "If W, then if D, then C" that expresses the bridge between warrant and the previous associated conditional.

Verheij (2005) argues that reasons against any of these statements can be seen as a kind of rebuttal of an argument that consists of warrant, data and claim (Fig. 3.2).

The first three are straightforward, and are clearly different. An argument against the datum that Harry was born in Bermuda (for instance by claiming that Harry was born in London) differs from an argument against the claim that Harry is a British subject (for instance by claiming that Harry has become a naturalized
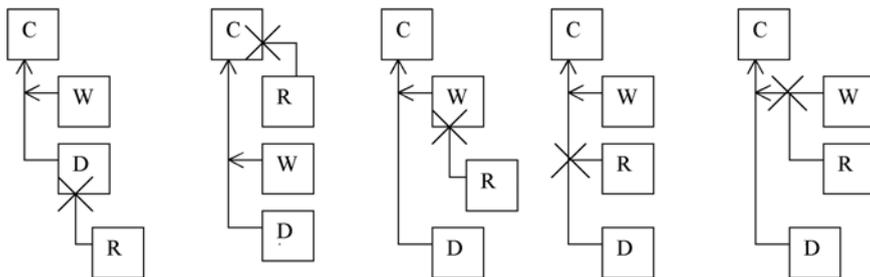
Fig. 3.2 Types of rebuttals (Verheij, 2005)

American) and from an argument against the warrant that a man born in Bermuda will generally be a British subject (for instance by claiming that those born in Bermuda are normally French). An argument against the fourth kind of statement (the first associated conditional) can be regarded as an attack on the connection between data and claim.

Such attacks have been characterized as "undercutting defeaters" by Pollock (1987). Harry having become a naturalized American could be an argument against the connection between Harry being born in Bermuda and Harry being a British subject. An argument against the fifth kind of statement can be regarded as an attack against the warrant's applicability: normally the warrant can justify the conditional that connects data and claim, but since there is a rebuttal, the warrant does not apply. In other words, when the associated conditional if W, then if D, then C, is not justified, the warrant, which normally gives rise to a bridge between data of type D and claim of type C, does not give rise to such a bridge for the actual data D and claim C at hand. For instance, Harry's parents both being aliens could well be an argument against the applicability of the warrant that a man born in Bermuda will generally be a British subject.

The three situations to which Toulmin attaches the term rebuttal (defeat of the conclusion, of the applicability of the warrant and of the authority of the warrant) are among these five kinds of rebuttals (the second, fifth and third, respectively). The other two kinds of rebuttals of a warrant–data–claim argument (i.e., the first and fourth kind) are apparently not mentioned by Toulmin (Verheij, 2005). Despite the limitations of Toulmin's framing of rebuttals, his outline of the role of rebuttals as well as the subsequent criticisms of his work in this respect have paved the way to establishing more dialogical patterns in classroom conversations in the science classroom. In our own work (Erduran et al., 2004) we designed a methodological approach where rebuttals were used as an indicator of improved reasoning (Table 3.1). Conversation with rebuttals, are, however, of better quality than those without given that individuals who engage in talk without rebuttals remain epistemologically unchallenged. The reasons for their belief are not questioned and are simply opposed by a counterclaim that may be more or less persuasive but is not a substantive challenge to the original claim. At its worst, such arguments are reducible simply to the enunciation of contrasting belief systems.

For instance, given that beliefs rely on justifications using data and warrants, a confrontation between a creationist and a Darwinist without any attempt to rebut the data or the warrants of the other would have no potential to change the ideas and thinking of either (Erduran et al., 2004). Only arguments, which rebut these components of argument can ever undermine the belief of another. Oppositional episodes without rebuttals, therefore, have the potential to continue forever with no change of mind or evaluation of the quality of the substance of an argument. Thus, arguments with rebuttals are an essential element of better quality arguments and demonstrate a higher level capability with argumentation. Furthermore, rebuttals can also be considered as a measure of conversational engagement. In other words, if one of the goals of promoting argumentation in science lessons is to engage learners in dialogical conversation where they can not only substantiate their claims but also refute others' with evidence, then the presence of rebuttals in conversation can act as an indicator of sustained engagement in argumentation discourse.

Contributions to Theoretical Representations and Frameworks

The analysis of arguments is often hard, not only for researchers but also for students and teachers. It is no surprise that for a technology-based task, one of my student teachers used a picture of me and sent me to space in an astronaut suit, labeling the photograph "in space, no-one can hear you argue!" A variety of tools and techniques have emerged from the theory of argumentation and critical thinking pedagogy that aim to help in the task of analysis (Reed & Rowe, 2005). Our work with pre-service science teachers (e.g., Erduran, 2006; Erduran et al., 2006) has led to the production of support tools including writing frames which help structure as well as evaluate arguments. One of the most common and intuitive of these tools is diagramming, by which the abstract form of an argument can be identified and seen at a glance, and according to which it is then possible to analyze more closely the relationships between an argument parts, for example Figs. 3.1 and 3.2. The utility of argument diagramming is seen in its almost universal adoption in the teaching of critical thinking and argumentation skills, as well as its deployment in various practical tools employed where complex argumentation is used as part of professional discourse. There is a wide range of diagramming techniques, some very general, some tailored to particular domains, for instance ARGUMED (Verheij, 2003a) and DEFLOG (Verheij, 2003b) systems.

A key technique used in various pedagogic and professional applications of argumentation theory is the "box-and-arrow" approach of identifying atomic components of an argument, and then indicating links between them with arrows. One of the first proponents of the approach in a pedagogic context was Beardsley (1950), and little has changed since then. In addition to identifying relationships of support between atoms in an argument, Reed and Rowe (2005) observe that the scheme has become refined to also identify four distinct ways in which compounds can be formed: as serial argument (in which one statement supports another, which

in turn supports a third); convergent argument (in which two or more statements independently support a third); linked argument (in which two or more statements jointly support a third) and divergent argument (in which two or more statements are supported by a third).

Complex argumentation (including verbal and written argumentation) can be constructed through arbitrarily complex combinations of these forms. Rather than viewing arguments as essentially just more or less complex binary relationships of support, Toulmin framed arguments as six-part complexes, comprising the Data, Warrant, Claim, Backing, Rebuttal, Qualifier. Though the starting point was jurisprudential, the resulting theory and its subsequent application are very general, and a Toulmin-style approach, including diagrams of argument components, is widespread in the literature including science education literature (e.g., Jiménez-Aleixandre et al., 2000; Simon et al., 2006). An important observation is that whatever the theoretical framework, be it Toulmin's or another author's, diagramming is much more than just ways of drawing pictures. Diagramming embodies many theoretical assumptions and conclusions, and works as a way of summarizing and applying substantial theories as practical tools that are simple and easy to understand.

Concluding Remarks

In this chapter I presented an overview of some methodological approaches in the study of argumentation in science classrooms. My review raises more questions than it provides answers. Some of the key challenges of qualitative research methods including the definition of the unit of analysis, and reliability and boundary markers within verbal as well as written data apply to argumentation analysis too. In this sense, the difficulties that science educators have experienced in applying Toulmin's framework to classroom conversations are not unique to Toulmin's framework as it is often claimed (e.g., Duschl, this book; Kelly & Takao, 2002). An analytical tool derived from whatever theoretical or grounded framework will have its limitations in application, and it will not answer many questions. The case of the difficulties researchers have experienced in the application of Toulmin's work in science education, in my view, is more representative of underspecification of the boundary markers that generate coding tools rather than an inherently limited feature of the framework itself. I have chosen to concentrate on Toulmin's work as an example to illustrate how his work has contributed to both the methodology and theory of knowledge in science education. My choice of Toulmin's framework is based on the observation that it has guided and influenced many researchers in the field. While other frameworks such as the work of Walton (1996) remain promising as methodological tools (Duschl, this book), there is not substantial work at the present time to warrant and attribute their contribution to methodology in the study of school science argumentation.

Many methodological challenges remain in addressing aspects of argumentation in the classroom that are understudied. In our work (Simon et al., 2006) we generated a preliminary typology for the classification of pedagogical strategies in the teaching of argumentation. Extension of our results to more definitive pedagogical models will necessitate the development of new tools of analysis of teaching. Extending the analysis of argumentation from verbal to more multimodal contexts where other representations (including gestures) can be regarded as components of argument also promises a fruitful territory for methodological studies. Likewise methodologies will need to be developed to be sensitive enough to capture issues at different levels of education including primary, secondary and tertiary students' and teachers' argumentation.

A significant gap in the literature concerns those aspects of the complex classroom environments including the sociological, political and psychological structures and processes that mediate argumentation in school science. Interdisciplinary investigations using science studies approaches (e.g., Duschl et al., 2006) promise a fruitful territory where new methodological approaches can be generated. It is noteworthy, however, that in the true spirit of argumentation, methodological questions will continue to challenge our understanding of teaching and learning processes thereby offering the potential to contribute to knowledge in science education.

References

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

Beardsley, M. C. (1950). Practical logic. Englewood Cliffs, NJ: Prentice-Hall.

Castells, M., Enciso, J., Cerveró, J. M., López, P., & Cabellos, M. (2007). What can we learn from a study of argumentation in the students' answers and group discussion to open physics' problems? In R. Pinto & D. Couso (Eds.), Contributions from science education research. Dordrecht, The Netherlands: Springer.

Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions that promote reasoning. Teachers College Record, 100(2), 315–368.

Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (in press). Analytic frameworks for assessing dialogic argumentation in online learning environments. Educational Psychology Review.

Duschl, R., Erduran, S., Grandy, R., & Rudolph, J. (2006). Guest editorial. Science Studies and Science Education, 90(6), 961–964.

Duschl, R., Ellenbogen, K., & Erduran, S. (1999). Understanding dialogic argumentation among middle school science students. Paper presented at the annual meeting of the American Educational Research Association, Montreal, April 1999.

Erduran, S. (in press). Breaking the law: Promoting domain-specificity in chemical education in the context of arguing about the Periodic Law. Foundations of Chemistry.

Erduran, S., & Dagher, Z. (2007). Exemplary teaching of argumentation: A case study of two middle school science teachers. In R. Pinto & D. Couso (Eds.), Contributions of Science Education Research. Dordrecht, The Netherlands: Springer

Erduran, S. (2006). Promoting ideas, evidence and argument in initial teacher training. School Science Review, 87(321), 45–50.

Erduran, S., Ardac, D., & Yakmaci-Guzel, B. (2006). Learning to teach argumentation: Case studies of pre-service secondary science teachers. Eurasia Journal of Mathematics, Science and Technology Education, 2(2), 1–14.

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

Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, F. (1996). Fundamentals of Argumentation Theory, Mahwah, NJ: Lawrence Erlbaum.

Govier, T. C. (1987). Problems in argument analysis and evaluation. Providence, RI: Foris.

Hample, D. (1977). The Toulmin model and the syllogism. Journal of the American Forensic Association, 14, 1–9.

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.

Jiménez-Aleixandre M.P., & Pereiro Muñoz, C. (2005). Argument construction and change when working on a real environmental problem. In K. Boersma, M. Goedhart, O. De Jong, & H. Eijkelhof (Eds.), Research and the quality of Science Education (pp. 419–431). Dordrecht, The Netherlands: Springer.

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

Kelly, G.J., Chen, C., & Crawford, T. (1998). Methodological considerations for studying science-in-the-making in educational settings. Research in Science Education, 28(1), 23–50.

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

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

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

Loui, R. P. (2005). A citation-based reflection on Toulmin and argument. Argumentation, 19, 259–266.

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. E. (1998). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin Press.

Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. International Journal of Science Education, 28(15), 1817–1841.

Marquez, C., Izquierdo, M., & Espinet, M. (2006). Multimodal science teachers' discourse in modeling the water cycle, Science Education, 90 (2), 202–226.

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.

Piccinini, C. L., & Martins, I. (2005). Comunicação Multimodal na sala de aula de Ciências. Ensaio Pesquisa em Educação em Ciências, 6(1), 1–14.

Pollock, J. L. (1987). Defeasible reasoning. Cognitive Science, 11, 481–518.

Reed, D., & Rowe, G. (2005). Translating Toulmin diagrams: Theory neutrality in argument representation. Argumentation, 19, 267–286.

Sampson, V., & Clark, D. B. (2006). The development and validation of the Nature of Science as Argument Questionnaire (NSAAQ). Paper presented at the National Association of Research in Science Teaching Conference, San Francisco, April.

Sandoval, W. A., & Millwood, K. A. (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.

Scerri, E. (2003). Philosophical confusion in chemical education research. Journal of Chemical Education, 80(5), 468–474.

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.

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.

Verheij, B. (2005). Evaluating arguments based on Toulmin's scheme. Argumentation, 19, 347–371.

Verheij, B. (2003a). Artificial argument assistants for defeasible argumentation. Artificial Intelligence, 150, 291–324.

Verheij, B. (2003b). DefLog: On the logical interpretation of prima facie justified assumptions. Journal of Logic and Computation, 13, 319–346.

Voss, J. (2005). Toulmin's model and the solving of ill-structured problems. Argumentation, 19, 321–329.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B.C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Ed.), The psychology of learning and motivation: Vol. 17, Advances in research and theory (pp. 165–213). New York: Academic Press.

Walton, D.N. (1996). Argumentation schemes for presumptive reasoning. Mahwah, NJ: Lawrence Erlbaum.

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, P. (2002). Scaffolding preservice science 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.