Enhancing the Quality of Argumentation in School Science

Jonathan Osborne,[1] Sibel Erduran,[2] Shirley Simon[3]

[1]*Department of Education and Professional Studies, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK*

[2]*Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol BS8 1JA, UK*

[3]*Institute of Education, University of London, London, UK*

Received 12 November 2003; Accepted 6 April 2004

Abstract: The research reported in this study focuses on the design and evaluation of learning environments that support the teaching and learning of argumentation in a scientific context. The research took place over 2 years, between 1999 and 2001, in junior high schools in the greater London area. The research was conducted in two phases. In phase 1, working with a group of 12 science teachers, the main emphasis was to develop sets of materials and strategies to support argumentation in the classroom, and to support and assess teachers' development with teaching argumentation. Data were collected by video- and audio-recording the teachers' attempts to implement these lessons at the beginning and end of the year. During this phase, analytical tools for evaluating the quality of argumentation were developed based on Toulmin's argument pattern. Analysis of the data shows that there was significant development in the majority of teachers use of argumentation across the year. Results indicate that the pattern of use of argumentation is teacher-specific, as is the nature of the change. In phase 2 of the project, the focus of this paper, teachers taught the experimental groups a minimum of nine lessons which involved socioscientific or scientific argumentation. In addition, these teachers taught similar lessons to a comparison group at the beginning and end of the year. The purpose of this research was to assess the progression in student capabilities with argumentation. For this purpose, data were collected from 33 lessons by video-taping two groups of four students in each class engaging in argumentation. Using a framework for evaluating the nature of the discourse and its quality developed from Toulmin's argument pattern, the findings show that there was improvement in the quality of students' argumentation. This research presents new methodological developments for work in this field. © 2004 Wiley Periodicals, Inc. J Res Sci Teach 41: 994–1020, 2004

Contract grant sponsor: UK Economic and Social Science Research Council; Contract grant number: R000237915.
Correspondence to: Jonathan Osborne; E-mail: jonathan.osborne@kcl.ac.uk
DOI 10.1002/tea.20035
Published online 2 November 2004 in Wiley InterScience (www.interscience.wiley.com).

Curriculum innovations in science, such those sponsored by the Nuffield Foundation in the UK and the National Science Foundation in the USA in the 1960s and 1970s, have had little impact on the practices of science teachers. Four decades after Joseph Schwab's argument that science should be taught as an ''enquiry into enquiry,'' and almost a century since John Dewey advocated classroom learning be a student-centered process of enquiry, we still find ourselves struggling to achieve such practices in the science classroom. Witness the publication of the AAAS edited volume on inquiry, the recent release of *Inquiry and the National Science Education Standards*, and the inclusion of ''scientific enquiry'' as a separate strand in the English and Welsh science national curriculum. These three works serve as signposts to an ideological commitment that teaching science needs to accomplish much more than simply detailing what we know. In addition to teaching the content, of growing importance is the need to educate our students and citizens about how we know and why we believe in the scientific worldview (Driver, Leach, Millar, & Scott, 1996; Millar & Osborne, 1998). Such a shift requires a new focus on: (1) how evidence is used in science for the construction of explanations—that is, on the arguments that form the links between data and the theories that science has constructed; and (2) the development of an understanding of the criteria used in science to evaluate evidence and construct explanations. Central to this perspective is a recognition that language is not merely an adjunct to science but an essential constitutive element (Norris & Phillips, 2003; Osborne, 2002). More specifically that the construction of argument, and its critical evaluation, is a core discursive activity of science.

Although the consideration of the important role language, conversation, and discussion have in science learning can be traced back three or four decades, it was not until the 1990s that serious discussion of the role of language in science learning began (see Lemke, 1990). However, a neglected but valuable body of initial work was conducted in the fields of literacy education, wherein the role of refutational texts in aiding student conceptual understanding was studied (Alvermann & Hynd, 1986; Hynd & Alvermann, 1986). Such texts, in addition to including the standard scientific explanation, also included arguments that refuted common misconceptions. The major finding of this body of work was that such texts help significantly with the development of students' scientific knowledge—a finding that has been confirmed by further studies conducted in the last decade (Guzetti, Synder, Glass, & Gamas, 1993; Hynd, Alvermann, & Qian, 1997; Hynd, McNish, Qian, Keith, & Lay, 1994). These findings suggest that addressing the epistemic basis of belief, a marginalized feature of traditional science education, is also likely to lead to more effective achievement of conceptual goals. An indication of the burgeoning interest in the field of language, literacy, and discourse in science education is the recent conference that drew together, for the first time, workers from both fields (Yore, Bisanz, & Hand, 2003).

Only more recently, however, has research in science education itself turned its attention to that discourse which specifically addresses argumentation. The general point is that argumentation—the coordination of evidence and theory to support or refute an explanatory conclusion, model, or prediction (Suppe, 1998)—is a critically important epistemic task and discourse process in science. Situating argumentation as a central element in the learning of sciences has two functions: one is as a heuristic to engage learners in the coordination of conceptual and epistemic goals, and the other is to make student scientific thinking and reasoning visible to enable formative assessment by teachers or instructors.

In addition, from a societal perspective, contemporary science impinges directly upon many aspects of people's lives. Individuals and societies have to make personal and ethical decisions about a range of socioscientific issues based on information available through the press and other media. Contemporary developments in science and technology (e.g., genetic engineering, reproductive technologies, food safety) often pose dilemmas for society, particularly where they are based on equivocal findings or contested claims whose resolution depends not simply on a knowledge of science but also on the application of moral and ethical values (Levinson & Turner, 2001; Ratcliffe & Grace, 2003). Evaluating media reports of such socioscientific issues is not straightforward as it requires the ability to assess whether the evidence is valid and reliable, to distinguish correlations from causes, observations from inferences, and to assess the degree of risk (Millar & Osborne, 1998; Monk & Osborne, 1997). Within the context of a society in which scientific issues increasingly dominate the cultural landscape, where social practices are constantly examined and reformed in light of scientific evidence, and where the public maintains an attitude of ambivalence (Giddens, 1990) or anxiety about science (Beck, 1992), there is therefore an urgent need to improve the quality of young people's understanding of the nature of argument in general, and argument in a scientific context in particular. Finally, it is ironic that science, which presents itself as the epitome of rationality, so singularly fails to educate its students about the epistemic basis of belief, relying instead on authoritative modes of discourse (Scott, 1998) that leave students with naive images of science (Driver, Leach, Millar, & Scott, 1996) and little justification for the knowledge they have acquired.

An important task for science education, therefore, is to expose the epistemic core of science—the use of argument to construct explanations of the material world and develop children's ability to understand and practice scientifically valid ways of arguing, enabling them to recognize not only the strengths of scientific argument, but also its *limitations* (Osborne & Young, 1998). The research presented herein examines whether the quality of "argument" of young peoples' argument about scientific issues could be enhanced in science lessons.

Previous Research on Argument

Over the past few decades, certain influential educational projects have laid foundations for the work on argumentation in science lessons. These projects have promoted independent thinking, the importance of discourse in education, and the significance of cooperative and collaborative group work (e.g, Barnes, 1976; Cowie & Ruddick, 1990; Ratcliffe, 1997; Rudduck, 1983; Solomon, 1990). In addition to these projects, a body of relatively unintegrated research concerning argumentative discourse in science education has begun to emerge (Alverman, Qian, & Hynd, 1995; Herrenkohl & Guerra, 1995; Herrenkohl, Palinscar, DeWater, & Kawasaki, 1999; Jiménex-Aleixandre, Rodríguez, & Duschl, 2000; Zohar & Nemet, 2002). Perhaps the most significant contribution to this literature came from Kuhn (1991), who explored the basic capacity of individuals to use reasoned argument. Kuhn investigated the responses of children and adults to questions concerning problematic social issues. She concluded that many children and adults (especially the less well educated) are very poor at coordinating and constructing a relationship between evidence (data) and theory (claim), which is essential to a valid argument. More recent work by Hogan and Maglienti (2001), exploring the differences between the reasoning ability of scientists, students, and nonscientists, also found that the performance of the latter two groups was significantly inferior to that of scientists.

Koslowski (1996), who was critical of Kuhn's emphasis on covariation, was less doubtful of young people's ability to reason, pointing to the fact that theory and data are both crucial to reasoning and interdependent. Hence, lack of knowledge of any relevant theory or concepts often constrains young people's ability to reason effectively. Although this is an important point (and one to which we will later return), Kuhn's research is significant because it highlights the fact that, for the overwhelming majority, the use of valid argument does not come naturally and is acquired through practice. The implication that we draw from the work of Kuhn and others is that argument is a form of discourse that needs to be appropriated by children and *explicitly taught* through suitable instruction, task structuring, and modeling. Just giving students scientific or controversial socioscientific issues to discuss is not sufficient to ensure the practice of valid argument. Similar conclusions were reached by Zohar and Nemet (2002) and Hogan and Maglienti (2001). The latter argued that "students need to participate over time in explicit discussions in the norms and criteria that underlie scientific work" (p. 683). Hence, for such reasons, the focus of our research was to develop pedagogical practices that support argumentation and foster students" epistemological development, because, although general advice concerning how to structure successful discussion and argumentation can be found in the literature (e.g., Dillon, 1994) or in other disciplines (Andrews, 1995), only a little has been situated within the specific context of the science classroom.

A significant problem confronting the development of argumentation in the science classroom is that it is fundamentally a dialogic event carried out among two or more individuals. Scott (1998), in a noteworthy review of the nature of classroom discourse, showed how it lies on a continuum from "authoritative," which is associated with closed questioning and IRE dialogue, to "dialogic," which is associated with extended student contributions and uncertainty. However, the combination of the power relationship that exists between a science teacher and student, the rhetorical project of the science teacher that seeks to establish the consensually agreed-upon scientific worldview with his or her students, and the authoritarian, dogmatic nature of the discipline means that opportunities for dialogic discourse are minimized. Deliberative discussions have commonly occupied only 2% of all science lessons in junior high schools (Lemke, 1990; Wells, 1999). Hence, introducing argumentation requires a shift in the normative nature of classroom discourse. However, change requires that science teachers be convinced that argumentation is an essential component of the learning of science. In addition, they require a range of pedagogical strategies that will both initiate and support argumentation if they are to adopt and integrate argumentation into the classroom.

At the core of such strategies is the requirement to consider not singular explanations of phenomena but *plural* accounts (Monk & Osborne, 1997). Students must, at the very least, spend time considering not only the scientific theory *but also an alternative*, such as the common lay misconception that all objects fall with the same acceleration versus the notion that heavier things fall faster. This approach shares the view developed in literacy studies that secure knowledge and understanding are as much a product of knowing why some ideas are erroneous as much as why other ideas are correct.

The evidence that does exist suggests that argumentation is fostered by a context in which student–student interaction is permitted and encouraged (Alexopoulou & Driver, 1997; Alverman et al., 1995; Herrenkohl et al., 1999; Jiménex-Aleixandre et al., 2000; Kuhn, Shaw, & Felton, 1997; Thorley & Treagust, 1986; Zohar & Nemet, 2002). For instance, Kuhn et al. (1997), in testing the hypothesis that engagement in thinking about a topic enhances the quality of reasoning about the issue, found that dyadic interaction significantly increased the quality of argumentative reasoning in both early adolescents and young adults, as did Alverman et al. (1995). Likewise, the work of Eichinger et al. (1991) and Herrenkohl et al. (1999) found that bringing scientific discourse to the classroom required the adoption of instructional designs that permit students to work collaboratively in problem-solving groups. Some of the research on discourse has also pointed to the importance of establishing procedural guidelines for the students (Herrenkohl et al., 1999). Hence, the message for our research was that both epistemological *and* social structures in the classrooms are important factors for designing activities that foster argumentation. From an epistemic perspective, there is the need to provide students access to not a singular worldview but to plural accounts of phenomena and the evidence that can be deployed in an argument. Focusing on the epistemic and social nature of classroom activities is an essential precursor to cognitive development, because, as Billig (1996) argued, ‘‘humans do not converse because they have inner thoughts to express, but they have thoughts because they are able to converse’’ (p. 141). Thus, learning to think is learning to argue.

Second, from a social perspective, there is the need to establish a social context that fosters dialogic discourse. This we have seen as the need to use techniques such as student presentations, small-group discussions coupled with guidelines and assistance that support the appropriation of argumentation skills and discourse. Consequently, in developing materials and strategies for argumentation we have used a consideration of the social and cognitive elements of classroom life as guiding principles that underlie the approach and design of all that we have sought to do.

Finally, at this point, it is worth noting that, in this study, we have drawn a distinction between ‘‘argument’’ and ‘‘argumentation.’’ The former we see as a referent to the claim, data, warrants, and backings that form the substance or *content* of an argument. The latter, in contrast, we see as a referent to the *process* of arguing. The focus of this work has been to explore those strategies that scaffold and support ‘‘argumentation,’’ and to develop frameworks for the assessment of its quality.

Research Objectives

Our analysis of the previous body of research led us to the belief that promoting the practice of ‘‘argumentation’’ in science lessons requires the development of appropriate pedagogical strategies and materials that offer practical guidance for teachers. Furthermore, the value and outcomes of such guidance need to be assessed through empirical studies. Therefore, our research sought to:

1. Identify some of the pedagogical strategies necessary to promote ‘‘argument’’ skills in young people in science lessons.
2. Trial the pedagogical strategies and determine the extent to which their implementation enhances teachers’ pedagogic practice with ‘‘argument.’’
3. Determine the extent to which lessons that follow these pedagogical strategies lead to enhanced quality in students’ arguments.

The focus of the study reported here is, however, principally on the third area of interest.

To investigate these objectives, we chose to work in two contexts—a socioscientific context and a scientific context. The former we saw as important because many of the debates surrounding science in the public domain are of this nature. Moreover, many such issues draw on existing knowledge and resources of which young children already have some knowledge. Scientific arguments are, however, important because they expose the justification for belief in the scientific worldview and the underlying rationality that lies at the heart of science.

Our approach to this research was to work initially, in the first year of our project, with a group of 12 junior high school teachers to explore and develop their practice at initiating argumentation in the classroom and then, in the second year, with a subset of 6 of these teachers to explore the effect such activities had on the classroom discourse and students’ use of argument.

Our Analytic Perspective Upon Argument

Assuming, as the research evidence suggests, that a context that fosters and develops students’ use of argumentation can be established, then what can teachers learn by listening to student discussion and how can they foster and improve the quality of argumentation? Essentially, how can they respond formatively to assist their students and develop their reasoning? How, for instance, can they identify the essential features of an argument? How are they to judge that one argument is better than another? Also, how should they model arguments of quality to their students? Before we could ask teachers to engage their students in argumentation and use the information they acquired from such an activity to plan subsequent lessons or evaluate students' learning, it was essential to provide some theoretical guidance to answer such questions. Thus, an important component of this research was the need to adopt and develop a set of criteria to analyze both the content and form of children's arguments.

In our work, we chose to use the analytic framework developed by Toulmin (1958). His model of argument, referred to here as Toulmin's argument pattern (TAP), was one of the first to challenge the "truth"-seeking role of argument and to consider, instead, the rhetorical elements of argumentation and their function. For Toulmin, the essential elements of argument are claims, data, warrants, and backings. At the base of all arguments is a claim—essentially an unwarranted assertion that a proposer believes has the status of a universal truth. Arguments, however, normally rely on evidence or justifications that consist of data related to the claim by a warrant. Warrants, in turn, may be dependent on a set of underlying theoretical presumptions or backings, which are often implicit. For instance, Galileo's claim that the heliocentric model of Copernicus was true was reliant on a set of data of observations of Jupiter's moons. Of themselves, such data do not establish the Copernican claim. Rather, these data require a warrant—that this observation falsifies a basic premise of the Ptolemaic system requiring all objects to orbit the Earth—which then justifies the claim. Arguments may also be hedged with qualifications to show the limits of the validity of the claim and are commonly rebutted by challenging either the data, warrants, or backings.

An alternative framework for the analysis of argument is that developed by Walton (1996), which characterizes argument in terms of a schema of 25 common forms of reasoning. Our view, however, is that Walton's framework gives more emphasis to the content of an argument, which was not the essential focus of our work. Toulmin's model, in contrast, places more emphasis on the generic features of argument, fitting better with our interest in argumentation in general. In addition, Toulmin's model has been used as a basis for characterizing argumentation in science lessons, and is implicit in a coding system of others (Kuhn et al., 1997; Pontecorvo, 1987; Schwarz & Lederman, 2002; Zohar & Nemet, 2002). Following these authors, we therefore used the Toulmin framework to focus on the epistemic and argumentative operations adopted by students—that is, their reasoning functions, and strategies. Features that we have concentrated on in developing our analysis of argumentation in both scientific and socioscientific contexts include the extent to which students have made use of data, claims, warrants, backings, and qualifiers to support their arguments and the extent to which they have engaged in claiming, elaborating, reinforcing, or opposing the arguments of each other.

The Research Program

General Features of the Research

A group of junior high school teachers interested in collaborating with us were initially established for some preliminary work in the area. From this group, 12 were selected—our principal criterion being selection of experienced and confident teachers of science, as the work would involve a degree of risk on their part requiring the use of innovative or unfamiliar pedagogy. All the teachers were located within schools in or around the greater London area. The data reported in this study were drawn from the classrooms of 6 of the 12 teachers. Three were working in inner-city urban comprehensives and one a suburban comprehensive—all of which had a highly diverse ethnic profile. The other two came from satellite towns of London and served a predominantly middle-class, monoethnic white community. One the six schools was an all-girls public school, one a private all-boys school, and the rest mixed public schools. This diverse range of schools meant that the research reported here was conducted with a sample of students of varying academic ability that is broadly representative of the range of ability found within the student population. Our discussions with teachers led to the choice of students in grade 8 (age 12–13 years) as the most suitable for the intervention because of the freedom from the curricular constraints imposed on teachers by public examinations.

The research was conducted in essentially two phases. In year 1, we sought to focus on developing the skills of the teachers and the materials for use in argument-based lessons. During year 1, the teachers attended six half-day meetings at King's College London, to discuss and share pedagogical strategies for teaching such lessons, to develop materials for teaching argument, and to develop their understanding of our theoretical perspective on argument. In both the first and second years, the teachers involved in the study incorporated a series of a minimum of nine argument-based lessons, approximately once per month over the course of 1 year, involving focused discussions relevant to the English national curriculum for science. The first and final lessons were devoted to discussion of a socioscientific issue of whether zoos should be permitted, whereas the remaining lessons were devoted solely to discussion and argument of a *scientific* nature. To support these lessons, teachers were initially provided with a set of materials drawn from the literature and our own ideas for use with students. These aimed to develop their knowledge and capabilities with scientific reasoning by examining evidence for or against a theory; for example, the particle hypothesis or the explanation of day and night. Other activities focused on sets of data, their interpretation, and conclusions that could be drawn from them. Resources for teaching all these lessons were also developed separately by teachers.

To assess the teachers' progress, we video- and audio-recorded the teacher at the beginning of year 1 and year 2, and systematically analyzed the transcripts for the components of argument identified by Toulmin to ascertain the teachers' use of argumentation and to measure their progress at argumentation. The outcomes of that work were evaluated by Simon, Erduran, and Osborne (manuscript in preparation).

In the second year of the project, the focus of this study, we worked with a subset of 6 of the 12 teachers and asked them to repeat the process they had undertaken in the previous year. These six teachers were selected because they were considered to have made more progress in their ability to facilitate and incorporate argumentation in their pedagogical practice. This was a judgment born out retrospectively by the data analysis of their classroom transcripts, with all teachers showing significant gains ($p < 0.05$) in the use of components of argumentation in classroom discourse.

Support in the second year was reduced to three half-day meetings throughout the year with feedback provided in situ whenever a visit was made for the purpose of data collection. In addition, each teacher taught a class of the same grade, as similar as possible in student aptitude and ability, to provide some basis for comparison of the effect of the intervention. In this phase, the focus of our analysis stage was on the recordings and transcripts of the discussions by students to determine whether there was any improvement in the quality or quantity of student argument. In what follows is a summary of the principal findings that have emerged from the work of the project and an exploration of their implications. In this paper, however, the principal focus is on the changes shown by the students.

Materials and Support for Argument

One feature of this work has been to try and develop materials that could be used for supporting argumentation in the classroom. The essential precursor to initiating argument in any context is the generation of differences, which in science is dependent on presenting alternative theoretical interpretations of any phenomenon. Hence, a common framework for most of the materials we have developed has taken the form of presenting or generating competing theories for students to examine, discuss, and evaluate. A universal requirement of the social structure of the lesson has been the opportunity for students to meet in small groups and discuss these ideas, to evaluate the evidence that does or does not support each theory, and to construct arguments justifying the case for one or other theory. However, as the work of Koslowksi (1996) has shown, initiating argument requires a resource or data to enable the construction of argument. Hence, commonly, competing theories have been accompanied by evidence that students are asked to use to decide whether the evidence presented supports theory 1, theory 2, both, or neither. Using these ideas, as the essential principles for initiating argument in the science classroom and drawing on the literature, we developed nine generic frameworks for promoting argument in the classroom. An outline of these is provided in Table 1 and Figure 1 gives one example in more detail.

The reasons for choosing to develop generic frameworks was essentially twofold. One was pragmatic in that the topics being taught by the teachers varied from school to school and lesson to lesson. Demanding that specific exemplar lessons be taught would have placed too restrictive a burden on the teachers of science and made the project unworkable. More fundamentally, providing a framework on which the substance of a lesson could be ‘‘hung’’ provided teachers with a vital element of independence. This enabled both of them to make a contribution in developing and trialing their own ideas, and to take ownership of the work—an element that is vital for successful curriculum innovation (Ogborn, 2002). Although a detailed analysis of the frameworks used in all the lessons has not been conducted, the majority of the materials developed have made use of framework 6 (Figure 1).

Another strategy was drawn from the literature on teaching students to write (Bereiter & Scardamalia, 1987). Constructing a good argument is not a simple task and students need guidance and support that will help them to ‘‘scaffold’’ and build their sense of what is an effective argument. Wray and Lewis (1997) have shown that when such genres of writing or expression are not familiar, ‘‘writing frames’’ that support the process of writing can provide vital support and clues as to what is needed. Essentially, these contain a set of stems such as: my argument is . . . ; my reasons are that . . . arguments against my idea might be that . . . ; I would convince somebody that does not believe me by . . . ; and the evidence to support my argument is . . . . These stems provide essential prompts necessary to initiate the construction of a written argument and to structure it in a coherent manner. Writing frames are then essentially drafting documents for recording notes of their discussion, which can then be used as a structure for producing a written argument. Several variants of these writing frames were developed for supporting argumentation.

Finally, an important aspect of developing an understanding of argument and evidence for students is the need to present examples of argument and model good practice. This involves offering students examples of both weaker and stronger arguments, enabling discussion of the features that make one better than another. Hence, exemplars of arguments of different quality that could be offered to students were produced and shown to the teachers engaged in the research. Examples of poor quality argumentation were written to illustrate that such arguments relied on assertion with minimal use of data or warrants to justify claims, whereas those developed to illustrate stronger argumentation drew on a wider range of evidence and included rebuttals of counter-arguments. Such exemplars serve an important illustrative function of what constitutes good argumentation—two of which are shown in what follows:

Table 1
Generic frameworks for materials for supporting and facilitating argumentation in the science classroom

Framework	Description
1. Table of statements	Students are given a table of statements on a particular science topic. They are asked to say if they agree or disagree with the statement and argue for their choices. This was developed from work on physical phenomena (Gilbert & Watts, 1983).
2. Concept map of student ideas	Students are given a concept map of statements derived from student conceptions of a science topic derived from the research literature. They are then asked to discuss the concepts and links individually and as a group to decide whether they are scientifically correct or false, providing reasons and arguments for their choice. This was an adaptation of the common use of concept mapping (Osborne, 1997).
3. A report of a science experiment undertaken by students	Students are given a record of another student's experiment and their conclusions. The experiment is written in a way to intentionally include information that is lacking or in a manner that could clearly be improved, so as to stimulate disagreement. Students are asked to provide answers to what they think the experiment and its conclusions could be improved, and why. This idea was drawn from the work of Goldsworthy, Watson and Wood-Robinson (2000).
4. Competing theories—cartoons	Students are presented with two or more competing theories in the form of a cartoon. They are asked to indicate the one they believe in and argue why they think they are correct. The work of Keogh and Naylor (Keogh & Naylor, 1999; Naylor & Keogh, 2000) has been valuable in developing an excellent stimulus for engaging children with scientific thinking.
5. Competing theories—story	Students are presented competing theories in the form of an engaging story reported in a newspaper. They are then asked to provide evidence supporting the theory they believe in and why.
6. Competing theories—ideas and evidence	In this approach, students are introduced to a physical phenomenon and then offered two or more, but generally two, competing explanations. In addition, a range of statements of evidence that may support one theory, the other, both, or neither are provided. In small groups, students are then asked to consider each piece of evidence and evaluate its role and significance. Finally, they must use the evidence to argue for one idea or another. This idea was adapted from the work of Solomon et al. (Solomon, 1991; Solomon, Duveen, & Scott, 1992).
7. Constructing an argument	Students are given an explanation of a physical phenomenon, such as day and night are caused by a spinning Earth, and a number of data statements (typically four). They then have to discuss which data statements provide the strongest explanation for the phenomenon and provide an argument why. This is an idea adapted from the innovative work of Garratt et al. (1999) in undergraduate chemistry.
8. Predicting, observing, and explaining	This activity, drawn from the work of White and Gunstone (1992), involves introducing a phenomenon to children without demonstrating it and asking students to discuss in small groups what they think will happen when the phenomenon is initiated, and justify their reasoning. The phenomenon is then demonstrated and, if what happens is the antithesis of that expected, students are then asked to reconsider and reevaluate their initial arguments. Discussion focuses on the theory that they advance for their prediction and the evidence to support it.
9. Designing an experiment	Students are asked to work in pairs to design an experiment to test a hypothesis, such as that a silver kettle cools faster. Their design needs to specify not only what variable should be measured but how often and what steps should be taken to ensure that the data obtained are reliable. Pairs then meet to discuss their design, to propose alternative procedures and to argue for their relative merits.

Example 1: Competing Theories (Framework 6)

Theory 1: Light rays travel from our eyes onto the objects and enable us to see them.
Theory 2: Light rays are produced by a source of light and reflect off objects into our eyes so we can see them.

Which of the following pieces of evidence supports Theory 1, Theory 2, both or neither. Discuss.

a. Light travels in straight lines
b. We can still see at night when there is no sun
c. Sunglasses are worn to protect our eyes
d. If there is no light we cannot see a thing
e. We 'stare at' people, 'look daggers' and 'catch people's eye'
f. You have to look at something to see it.

Figure 1. An example of materials developed by one teacher using framework 6.

Weak argument:

We must see because light enters the eye [claim]. You need light to see by [data]. After all, otherwise we would be able to see in the dark [warrant].

Stronger argument:

Seeing because light enters the eye makes more sense [claim]. We can't see when there is no light at all [data]. If something was coming out of our eyes, we should always be able to see even in the pitch dark [rebuttal]. Sunglasses stop something coming in, not something going out [data]. The only reason you have to look towards something to see it is because you need to catch the light coming from that direction [rebuttal]. The eye is rather like a camera with a light-sensitive coating at the back, which picks up light coming in, not something going out [warrant].

Implementing the social dimension of the work required teachers to employ a set of social structures that permitted and encouraged dialogic interactions between student and student, and student and teacher. Foremost was the need to provide opportunities for small-group discussions and guidance on how students should interact. For instance, the teachers were asked to explain to their students the importance of thinking of counter-arguments that challenge the justification of another's argument. In addition, we sought to devise structures by which student argumentation could be facilitated and scaffolded through the use of a set of argumentation prompts. Essentially these were open-ended questions designed to elicit a justificatory argument from the student, such as: Why do you think that? Can you think of another argument for your view? Can you think of an argument against your view? How do you know? What is the evidence for your view? These and other strategies were very much the focus of the sessions held with the teachers during the first year, where ideas and approaches were shared and expertise at argumentation constructed collaboratively.

Data Sources

In this phase of our work, six teachers worked with us teaching argument in the classroom. Using the data gathered from their students we sought to examine the development in students' ability to incorporate and use argumentation in two ways:

1. In a *socioscientific context*, by comparing the data from the experimental group with that from a comparison group who were also taught the same zoo lesson by the same teacher at the beginning of the academic year. At the end of the year, the same teacher taught a similar lesson about the possible siting of a leisure center in a nature reserve to both the experimental and comparison groups. These data enabled comparison of the gains achieved by the control and comparison group in the course of a year.
2. In a *scientific context*, by comparing the development of the experimental group using data from an argumentation lesson taught at the beginning of the academic year and a different lesson at the end of the year. These data were used to identify the nature and development of students' argumentation in a scientific context throughout the course of a year. By comparing these data with those obtained from argument in a socioscientific context, it was also possible to compare the quality of argumentation in the two contexts.

Teachers began the second year with a student task they had used in the first phase of the work argumentation in a socioscientific context—an exploration of arguments for, and against, the funding of a new zoo. Each lesson had three sections. At the onset, the teacher distributed a letter outlining the task, with an ensuing whole-class discussion on the pros and cons of zoos. The students were then placed into small groups of three or four and asked to discuss whether or not the zoo should be built. Finally, in the last phase of the lesson, the groups made presentations and shared their opinions with the rest of the class. For homework, students were typically asked to write a letter or to compose a poster that would communicate their arguments. Needless to say, there was considerable variation between teachers in the detail of their implementation. Microphones were attached to the teachers so as to capture their verbal contribution to the lesson as well as their interactions with students during the group format.

Students were selected for the groups to be recorded by the teacher on the basis that their ability levels diverged somewhat but not excessively. Research has shown that groups of this composition are most likely to be functionally effective (Kutnick, Blatchford, & Baines, 2002). Groups containing students who had previously demonstrated considerable reticence at verbalizing their thinking were not chosen for the study and no attempt was made by the research team to control the mean level of ability groups that varied from school to school and teacher to teacher. Rather, the range of schools used for the study led naturally to considerable variation in the sample of students, which forms the basis of this study. Student discussions were recorded with a video camera with a fixed microphone placed in the center of the group. The same groups of students were used for the second phase of data collection at the end of the academic year. Inevitably, with absence and student mobility, there was minor attrition among the groups.

Thus, data sources for the findings reported here were: (a) the video and audio transcripts of the discussion from two groups of four grade 8 (age 12–13 years) students in each class (6 teacher tapes, 11 student videos) exploring arguments for and against the establishment of a new zoo at the beginning of the academic year. Early in the year, in addition to the lesson requiring argumentation about the merits and demerits of a zoo, these teachers also taught a lesson requiring argumentation in a scientific context, wherein data were collected by video recording the discussion of the *same* groups and audio recording the teacher's discourse (6 teacher tapes, 12 student videos)—set (b). Also at this time, data [set (c)] were collected from the same teachers teaching the same lesson to a comparison group (5 teacher tapes, 9 student videos) from the same grade. It was logistically impossible to control the nature of the comparison group, but teachers were asked to arrange, with a colleague, for the opportunity to take a class for these argumentation lessons with a group whom they judged to be similar in ability. Data from the comparison group, collected at the beginning and the end of the interventions, were used as a means for evaluating its overall effect.

In the intervening period, teachers taught a minimum of eight lessons using argument*in a scientific context*, which formed the substance of the intervention. This was the minimum number of argument lessons we asked of the teachers—and all were able to meet this requirement. The students in the comparison group undertook their normal scheme of work and did not engage in any lessons that explicitly required argumentation.

At the end of the year, another set of data was collected from the same group of 6 teachers teaching argumentation to the intervention class in a socioscientific context [set (d)—6 teacher tapes, 12 student videos] and in a scientific context [set (e)—5 teacher tapes, 10 student videos]. Again, data were collected by audio-taping the teachers and video-taping the same set of four students[1] whenever possible. In addition, a final set of data [set (f)] was collected from the comparison group for argumentation in a socioscientific context with an identical lesson (5 teacher tapes, 10 student videos). Field notes were also collected of salient features of all lessons and the materials used by the teachers. Table 2 provides a summary of all the data sets collected.

Finally, a semistructured interview was also conducted with the teachers at the beginning of each year to ascertain their views on argumentation. These interviews sought to identify teachers' perceptions of the salience of teaching argumentation to students and their understanding of its significance. Such interviews were also used as a means of identifying any changes that had occurred over the year. Each interview was recorded and transcribed. The interviews included questions on how teachers felt about their zoo lesson and what they viewed as important for student participation and learning of argumentation. However, because the focus of this paper is on the change in the quality of students' argumentation, these data are not reported here. Rather, using these data, we sought to answer our third research question by examining the students' development with argumentation.

Analyses: Assessing Students' Development With Argumentation

All audio-tapes were transcribed and analyzed to determine the nature of argumentation in whole-class and in small-group student discussion formats. Toulmin's (1958) model of argument was used as an analytical framework to identify the salient features of argument in the speech. The following section illustrates our method of coding the transcripts using TAP as a guiding framework. Fuller details can be found in Erduran, Osborne & Simon (in press). In the case of the following example of student discourse:

Table 2
Summary of data sets collected of argumentation

Beginning of year	End of the year
(a) Experimental group: Socioscientific argumentation activity about the siting of a zoo (6 teacher tapes, 11 student videos[a])	(d) Experimental group: Socioscientific argumentation activity about the siting of a leisure center (6 teacher tapes, 12 student videos)
(b) Experimental groups: Argumentation in a scientific context (6 teacher tapes, 12 student videos)	(e) Experimental groups: Argumentation in a scientific context (5 teacher tapes, 10 student videos)
(c) Comparison group: Socioscientific argumentation activity about the siting of a zoo taught by same teachers (5 teacher tapes, 9 student videos)	(f) Comparison group: Socioscientific argumentation activity about the siting of a leisure center taught by same teachers (5 teacher tapes, 9 student videos)

[a]Due to a set of factors, such as changes in teachers' timetable and occasional technical problems, a complete data set does not exist for all lessons.

Zoos are horrible, I am totally against zoos.

our focus was on the substantive claim. In this case, the difficulty lies in the fact that both can be considered to be claims; that is:

Zoos are horrible and *I am totally against zoos.*

The question for the analysis then becomes deciphering which of these is the substantive claim and which is a subsidiary claim. Our general view is that there is inevitably a process of interpretation to be made and that some of that process relies on listening to the tape and hearing the force of the various statements. Part of this might be substantiated by Austin and Urmson's (1976) distinction between locutionary statements—ones that have an explicit meaning, and perlocutionary statements—ones that have implicit meaning. The perlocutionary force with which these statements are made—something that can often only be determined by listening to the tape—is an aid to resolving which statement is intended as the substantive claim and the locutionary meaning.

Thus, our approach to the work was always to seek to identify, through either a careful reading of the transcript or, alternatively, listening to the tape, what constituted the claim. Once, the claim was established, the next step was resolution of the data, warrants, and backings. Our view here was that a necessary requirement of all arguments that transcend mere claims is that they are substantiated by data. Therefore, the next task was the identification of what constitutes the data for the argument, which is often preceded by words such as "because," "since," or "as." The warrant, if present, was then the phrase or substance of the discourse that relates the data to the claim.

Nevertheless, in undertaking this task, we were conscious of the methodological difficulties in using TAP as a method of determining the structure and components of an argument (Kelly, Drucker, & Chen, 1998). Reducing these difficulties was, therefore, a significant methodological challenge for our work. Hence, the procedure that we adopted was to make a distinction between what we see as *first-order* elements of an argument—that is, claims, grounds, and rebuttals—and *second-order* elements, which are the components of the grounds for the claim—that is, the data, warrants, and backings. The advantage of using such a schema is that it circumvents the main methodological difficulty of Toulmin's framework—the resolution of the second-order components.

Transcripts were examined and student discourse categorized into one of four categories: teacher talk; student talk that advanced claims only; student talk that consisted of claims *and* grounds (either data or warrants); and student talk that was non-argumentative and of a procedural or off-task nature. A sample of such talk with its coding is as follows:

Teacher:	Okay, so you are saying that if the moon is light the light is fire and fire needs oxygen. All right. That's kind of added to the stuff . . . you have actually talked about . . . but what Mark was saying about the shadow and light, the moon passes through a shape, it is in the shadow from the earth, and you can't see it. So we know it doesn't give out light. B, the moon shrinks. Let's discuss this one. Michael . . . why is the moon.	*Teacher talk*
Student:	The moon is solid and it can't expand.	*Student claim*
Teacher:	It can't expand. What were you saying about water?	*Teacher talk*
Student:	It can't expand because it hasn't got water on it.	*Student claim with grounds*

Table 3
Percentages of group discourse of an argumentative nature

Type of discourse	Lesson					
	Zoo lesson (%)	Zoo comparison group (%)	Science 1 (%)	Science 2 (%)	Leisure center (%)	Leisure center comparison (%)
Claims	4	3	4	5	4	3
Grounds	28	24	11	12	22	26
Non-argument	8	8	15	9	13	10
Teacher	59	64	71	75	61	61

Counts of the number of words uttered were then made and the results for all the tapes are shown in Table 3 for the elements of argumentation. The table shows the type of discourse in each type of lesson: the zoo lessons and Science 1 lessons that took place at the beginning of the year, and the Science 2 and leisure center lessons at the end of the year.

The data illustrate several features of the nature of the discourse in these lessons. First, earlier research has shown that deliberative discourse of a dialogic nature commonly occupies 2% or less of all classroom discourse in normal science lessons. These data, however, show that, in these lessons, argumentative discourse (claims, claims + grounds) now occupies 15–32% (sum of "claims" + "grounds" for Science 1 lesson) of the total discourse, which represents a major shift away from the normative form of authoritarian dialogue that permeates science classrooms. What is important about this finding is that it shows that it *is possible* to transform the nature of the discourse of science classrooms to one that is more genuinely dialogic rather than one that has been characterized as the last authoritative sociointellectual discourse remaining on the school curriculum (Ravetz, 2002).

The second notable feature of these data is that argumentative discourse is significantly less for argumentation in science lessons than it is for socioscientific lessons, suggesting that initiating argument in a scientific context is harder and more demanding both for students and their teachers, whose responsibility it is to scaffold such discourse. This finding emphasizes Koslowski's (1996) point that knowledge of the evidence is a necessary requirement to engage in reasoning and argumentation about phenomena.

Third, these data also show that there is little difference in the amount of discourse between the experimental groups and the comparison groups, suggesting that the amount of argumentative discourse is a feature of the teachers' structuring and organization of the lesson rather than any characteristic of the student groups. In short, the epistemic and social structures of the classroom are major determinants of the nature of the classroom discourse.

Assessing the Quality of Argumentation

To assess the quality of students' argumentation, transcripts were then searched to identify genuine episodes of oppositional analysis and dialogic argument. Opposition took many different forms and many arguments were coconstructed wherein students provided data or warrants for others' claims. Transcripts of group discussions (two groups per teacher) were examined to determine the number of episodes of explicit opposition in student discourse. In other words, the instances whereby students were clearly opposed to each other were traced. Typically, these instances were identified through the use of words or phrases such as "but," "I disagree with you," and "I don't think so." Once these episodes were characterized in the group format, they were reexamined for the interactions among the students in terms of who was opposing whom and who was elaborating on what idea or reinforcing and repeating an idea. In this fashion, the pattern of interaction for each oppositional episode was recorded for two groups from each teacher's classroom. The main processes identified in such episodes were opposing claims by other (O), elaboration (E), or reinforcement (R) of a claim with additional data and/or warrants, advancing claims (C) or adding qualifications (Q) (see Example 1 in what follows). Such analysis helps to identify the features of the interaction and the nature of the engagement between the students.

The Nature of Opposition. Each oppositional episode was analyzed using TAP to identify the principal components of an argument being deployed by the individuals in the group. In these episodes, claims were not always clearly stated but rather implied or extracted through questioning. All episodes were read independently by two coders who then met to compare their analysis and resolve differences in interpretation. Assessments of reliability conducted obtained agreement for different episodes in excess of 80%. These oppositional episodes are characterized by a diverse range of arguments and some examples are provided later to illustrate the nature of our analysis and the results.

The essential issue raised by these episodes relates to how to define their quality. What, for instance, makes one better than another? To answer this question, we developed a framework for the analysis of quality (Table 4). In establishing this framework, we drew two major distinctions. The first asks whether an argument contains any reasons and grounds (i.e., data), warrants, or backing to substantiate its claim as transcending mere opinion, and developing rational thought is reliant on the ability to justify and defend one's beliefs. Hence, we see the simplest arguments as those consisting of a claim. Some investigators, such as Zohar and Nemet (2002), would not wish to recognize claims without justifications as meriting any significance. However, we believe they are important because they are the first step toward initiating the process of establishing difference. Although we recognize that the opposition may simply consist of a counterclaim, which is essentially a discursive interaction incapable of any resolution, such moves do permit establishment of difference and higher quality argumentation. In addition, teachers need to be able to identify such discourse moves and expose their limitations—the lack of justification—to their students. Hence, our second level is arguments accompanied by grounds containing data or warrants,[2] followed by arguments consisting of claims, data, warrants, and rebuttals.

Episodes with rebuttals are, however, of better quality than those without, because oppositional episodes without rebuttals have the potential to continue forever with no change of mind or evaluation of the quality of the substance of an argument. Moreover, as Kuhn (1991) argued, the ability to use rebuttals is "the most complex skill," as an individual must "integrate an

Table 4
Analytical framework used in for assessing the quality of argumentation

Level 1:	Level 1 argumentation consists of arguments that are a simple claim versus a counterclaim or a claim versus claim.
Level 2:	Level 2 argumentation has arguments consisting of claims 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 as well, but this is not necessary.
Level 5:	Level 5 argumentation displays an extended argument with more than one rebuttal.

original and alternative theory, arguing that the original theory is more correct'' (p. 145). Thus, rebuttals are an essential element of arguments of better quality and demonstrate a higher-level capability with argumentation. This analysis led us to define quality in terms of a set of five levels of argumentation (Table 4).

The two examples that follow are provided to illustrate how our analysis has been applied to the data.

Episodes Without Rebuttals

Example 1. In this example, taken from the zoo lesson, a claim is advanced supported by some data:

O	S1:	I don't think they would hurt them in a professional zoo.
	S2:	But they might scare the other animals by seeing some sedated animal being dragged off.
O	S1:	Maybe stress.
	S3:	Not stress. Distress.

Here, what we have is a claim that professional zoos would not hurt animals, which is countered by claim that animals in zoos might be scared (claim) as they would see other sedated animals being dragged off (data). Thus, our summary of this example is that it consists of:

Claim versus counterclaim + data

Moreover, despite some embedded complexity, as an example of arguing we would contend that it is essentially weak because there is no attempt at a rebuttal (by either party), permitting the justification of belief by both parties to remain unexamined. Therefore, we would consider this to be argumentation at level 2.

Episodes With Rebuttals

Our essential distinction here is between episodes with weak rebuttals—that is, counter-arguments that are only tenuously related to the initial claim (level 3), episodes with a single rebuttal (level 4), and episodes with multiple rebuttals (level 5). Example 2 illustrates a case of a weak rebuttal, whereas Example 3 a clear, unambiguous rebuttal.

Example 2. The episode beneath begins with the implicit claim that zoos are beneficial. The data for this argument are that ''some animals wouldn't be able to breed in the wild'' and there is a warrant supplied that this is because ''they may not have enough food.'' This claim is further supported or elaborated by the claim that ''the animals need a safe place to live'' and the data to support this claim are that otherwise ''they will be at risk from predators.'' This second claim is weakly rebutted with a negation that is thinly supported by the data that the risk from predators is just ''nature.'' However, as the rebuttal of the proponent's data does not make a clear, self-evident connection to the data supporting the original claim, we consider this to be an example of a weak rebuttal and a level 3 argumentation. A summary of this argument would be that it consists of:

Claim (+ data + warrant) + claim (+ data) versus weak rebuttal (+ data)

O
S1: Some animals wouldn't be able to breed in the wild, because they may not have enough food.
S2: No, no, no, because an animal.........
E
S3: Extinction.
S1: The animal needs a place to live because they would be at risk from other predators.
E
S2: What are you putting?
S1: A place to live, or they would be at risk from other predators.
O
S3: They might not have enough food to eat.
O
S2: But I mean, that's nature, one has to....
S1: But we are for it

Example 3. Our third example is an argument taken from a scientific context where students have been given alternative theories to explain the phases of the moons that are on a numbered card, A, B, C, or D, referred to in the dialogue:

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

Here, the first student advances the claim that it is explanation A, appealing to a datum that "the part of the moon that gives out light in not always facing us." There is then a rebuttal supplied with supporting data that the "light that comes from the moon is actually from the sun" and a warrant that is unfinished.

Our summary of this argument would be that it consists of:

Claim (+ data) versus rebuttal (+ data + warrant)

This schema of analysis enabled us to analyze all transcripts for the level of argumentation achieved and to make various comparisons of the performance of the different groups at argumentation. In total, we identified 183 oppositional episodes from 63 group discussions in 33 lessons for all the argument lessons with the experimental and comparison groups. On average, there were approximately 3 oppositional episodes per group per lesson. Figure 2 shows the distribution of arguments by level for all of the oppositional episodes obtained from 135 episodes in 43 discussion groups in 23 lessons from the experimental group.

Figure 2 shows that the largest number of arguments emerging from the data, at both the beginning and the end of the year, was at level 2 (38% and 30%, respectively). Encouragingly, however, whereas at the beginning of the year only 40% of student arguments were at level 3 or above, by the end of the year the corresponding figure was 55%. Although this change was not significant, it does show a positive development in the quality of argument. Moreover, the number of level 1 arguments decreased from 22% to 15%. This finding is particularly encouraging because

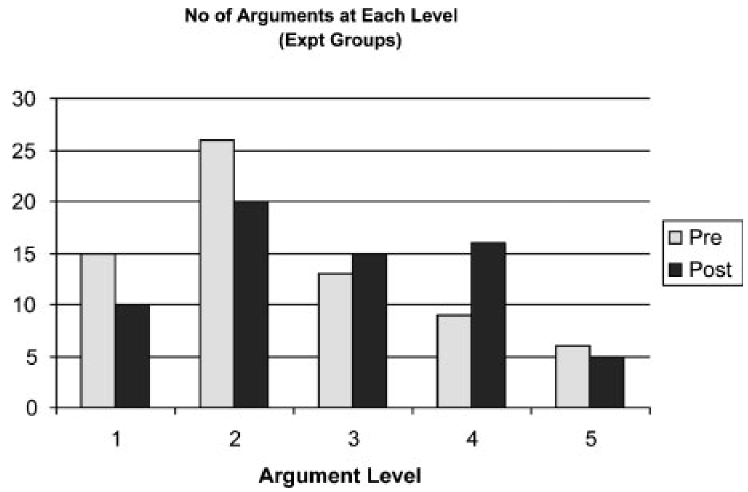

Figure 2. Chart showing numbers of each level of argumentation achieved in each oppositional episode ($n = 135$).

it suggests that only a small minority of arguments developed by students did not attempt to offer a rationale or some grounds for their claims, and that the intervention led to a decrease in the number of such arguments. Pedagogically, a preponderance of level 1 arguments would be problematic in that it is these types of argument that have the most potential for argumentation, which reinforces the lay perception of ‘‘argument as war’’ (Cohen, 1995). Rather, the metaphor of argument we chose to use in our work was of argument as a process of collaborative brainstorming toward the establishment of ‘‘truth’’ or better understanding—the primary goal of science. This view was succinctly summarized by Bachelard (1940) in his statement that ‘‘two people must first contradict each other if they really wish to understand each other. Truth is the child of argument, not of fond affinity.’’

This method of analysis permits a number of comparisons of the performance of the groups. First, it is possible to compare the distribution of levels achieved by the experimental group, at the beginning of the year in the first zoo lesson and their first science lesson, with those achieved at the end of the year in the last science lesson and their final leisure center lesson. This analysis shows that there was a shift toward the end of the intervention to more arguments of higher quality, shown more clearly by Figure 3. However, this shift was not significant.

Likewise, Table 5 shows a comparison of the levels of argument achieved by the groups in the discussion about the merits of zoos in the first zoo lesson with that 10 months later about whether a leisure center should be placed in an area of well-established wildlife.

The difference between these two distributions was not significant, although the pattern again suggests that there were more high-quality arguments at the end of the intervention than at the beginning.

It is also possible to compare the levels of argument of a scientific context achieved at the beginning versus at the end of the year Figure 4. Again, this shows that there was an improvement. Whereas, at the beginning of the year, 74% of the arguments were at level 2 or lower, at the end of the year this figure decreased to 50%.

One of the features of interest in this work was how the context of argument (i.e., scientific or socioscientific) affected the quality of argument. A comparison of the levels of argumentation achieved in the socioscientific lessons (zoo and leisure center) with those achieved in the science

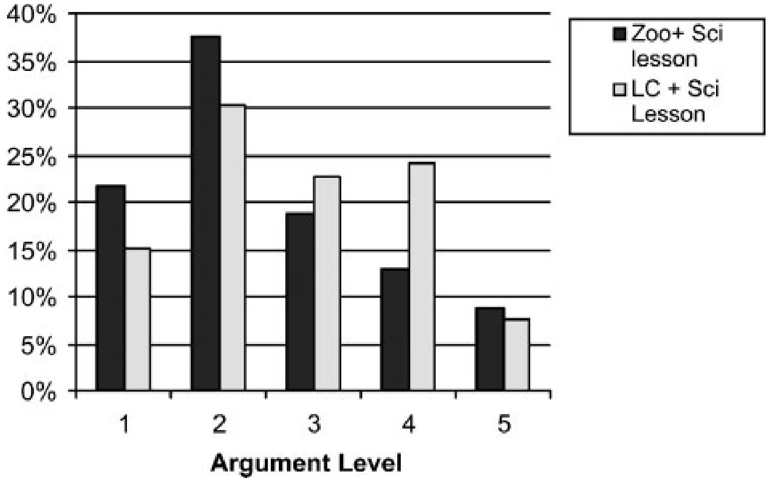

Figure 3. Levels of argumentation achieved by experimental groups, pre- ($n = 69$) and postintervention ($n = 66$).

lessons (Table 6) indicates that, in general, higher levels of argument were achieved in a socioscientific context and that this difference was significant ($p < 0.05$).

Taken together, with our analysis (Table 3) of the discourse in lessons showing that there was substantively less argumentative discourse in science lessons, these findings suggest that it is harder to initiate argumentation and argument in a scientific context than in a socioscientific context. However, whether the quality of argumentation is dependent on the quantity of argumentative discourse remains an open question.

Another feature of the research design was the use of a set of classes for comparison. In addition to teaching the lessons to the treatment group, we asked each of the teachers to teach the same zoo lesson to a class with similar academic ability at the beginning of the year. Likewise, at the end of the year, we asked the teachers to teach the leisure center lesson to both the experimental class and the comparison group. This enabled the performance of the two groups at the beginning of the year and at the end of the year (Table 7) to be compared. Although not a strict control in that it was pragmatically impossible to constrain variation in the groups, the contrast between these groups and the experimental groups does offer some insight into the outcomes of the intervention.

The data suggest that there was no significant difference between the groups at either the beginning or the end of the year. This finding suggests that, although the experimental group

Table 5
Levels of argumentation (socioscientific context) achieved at the beginning of the year (zoo lesson) and at the end of the year (leisure center lesson)

	Argument level achieved				
Lesson	1	2	3	4	5
Zoo Experiment (6 lessons, 11 groups)	13% (5)	34% (13)	21% (8)	16% (6)	16% (6)
Leisure Center Experiment (6 lessons, 12 groups)	11% (4)	32% (12)	16% (6)	32% (12)	11% (4)

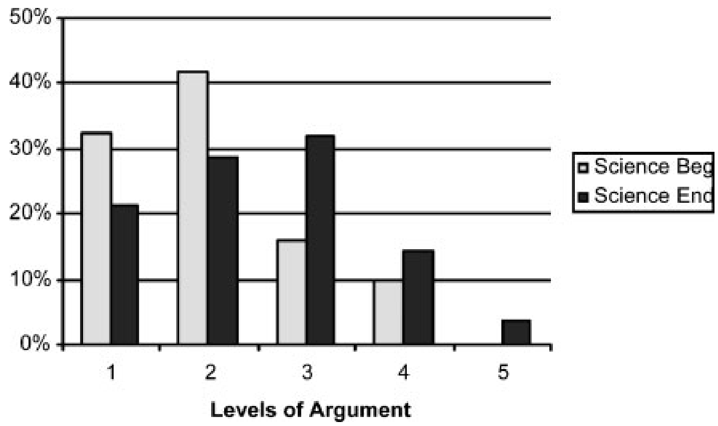

Figure 4. A comparison of levels of argumentation achieved by experimental groups in a scientific context at the beginning of the year ($n = 31$) versus that achieved at the end of the year ($n = 28$).

showed improvement in quality of argumentation, the comparison group also appeared to improve. However, a number of caveats must be placed on any interpretations of these data. First, the sample size was very small; second, the similarity between the groups remains questionable, because, apart from asking for the teacher to select a class with students of similar ability, it was impossible to impose any other constraints or control other variables that would have enabled a effective comparison between the two groups, such as for gender or ethnic mix. Third, it would be unrealistic to place too much emphasis on data collected from a very limited number of lessons at the end of the year.

If, however, these comparisons are valid measures, then there are two hypotheses as to why both groups have improved and why no difference between the experimental and comparison group was found: (a) that the improvement represented a natural developmental growth in individuals' reasoning and linguistic capability; or (b) that the improvement was a reflection of the individual teacher's growing ability to structure and facilitate argumentation. Of these two hypotheses, the extant evidence that suggests little growth in students' reasoning ability in adolescence (Hogan & Maglienti, 2001; Kuhn, 1991) indicates that the former is unlikely and that, instead, the improvement with the comparison group was a reflection of the development in the teacher's ability to foster a context for argumentation in the classroom. Our view, however, is that there are insufficient data here to resolve this question and that, with hindsight, such methodological approaches have little value unless they are undertaken with considerably larger

Table 6
Comparison of levels of argumentation (socioscientific context) with those achieved in a scientific context

Lesson	Argument level achieved				
	1	2	3	4	5
Zoo experiment & LC experiment (12 lessons, 23 groups)	12% (9)	33% (25)	18% (14)	24% (18)	13% (10)
Science lessons (11 lessons, 22 groups)	27% (16)	36% (21)	24% (14)	12% (7)	2% (1)

LC, leisure center.

Table 7
Comparison of levels achieved by experimental groups with those achieved by comparison groups

Lesson	Argument level achieved				
	1	2	3	4	5
Beginning of year					
Zoo experiment (6 lessons, 11 groups)	13% (5)	34% (13)	21% (8)	16% (6)	6% (6)
Zoo comparison group (5 lessons, 9 groups)	18% (5)	46% (13)	25% (7)	11% (3)	0% (0)
End of year					
Leisure center experiment (6 lessons, 11 groups)	11% (4)	32% (12)	16% (6)	32% (12)	11% (4)
Leisure center comparison group (5 lessons, 9 groups)	18.5% (5)	18.5% (5)	25.9% (7)	18.5% (5)	18.5% (5)

sample sizes. Thus, we point to the fact that what we have attempted here is essentially a "design experiment" (Brown, 1992), where the "learning effects are not even simple interactions, but highly dependent outcomes of complex and social and cognitive intervention"—all of which have the potential to confound the data and their interpretation. More significantly, we believe that this intervention achieved the positive effects we desired, albeit not to the extent we had hoped. In this sense, our intervention found that such treatments are capable of improving young children's quality of argumentation.

Discussion and Conclusions

In this study we have presented the major findings emerging from our work on developing argumentation in school science classrooms, its analysis, and the assessment of its quality. Methodologically, we believe our work has made progress on several fronts. First, the work sought to develop with teachers sets of materials that can be used in a structured and focused manner to facilitate argumentation in the classroom. As a result, we believe that we have gained some insights into the means of establishing a context that facilitates argumentation in the classroom, both in terms of the materials and the pedagogic strategies required for its support.

Second, our work with teachers has led to a change in the practice of the majority of this group, leading us to believe that, despite the many obstacles and barriers posed by the demands to implement different and innovative practice, it is possible for science teachers to adapt, change, and develop their practice into one in which there is a fundamental change in the nature of classroom discourse. One of the biggest fears expressed initially about this kind of work by some of the teachers was that the presentation of plural explanatory theories would confuse the children or lead to the development or strengthening of a belief in a scientifically incorrect idea. Such a reaction is comprehensible because the rhetorical project for the science teacher is to present a carefully crafted and persuasive argument for the scientific worldview (Osborne, 2001). Presenting alternatives to the scientific explanation would initially seem to undermine that project and naturally generate hesitation and doubt, if not resistance, in teachers. Yet, it was notable in the interviews at the end of the project that this initial concern was much diminished, if not absent altogether. In short, the teachers had come to recognize that the opportunity for students to reflect, discuss, and argue how the evidence did or did not support the theoretical explanation made debating the scientific case after the argumentation lesson a much simpler task, and one with which students were already engaged.

Third, one of the many problems that bedevils work in this field is a reliable systematic methodology for (a) identifying argument and (b) assessing quality. Our adoption and adaptation of Toulmin's argumentation pattern provided us with a method for discriminating the salient features of argumentation—the claims, rebuttals, and justifications—which are critical for developing and evaluating practice with argumentation in the classroom. This is not to say that the full Toulmin framework is of no value. Currently, at least in the UK, the language used to describe the epistemic components of science is that of the ''ideas'' of science and their supporting ''evidence.'' ''Ideas,'' on the one hand, consist of hypotheses, theories, and predictions that are essentially claims, whereas the data, warrants, backings, rebuttals, and qualifiers are the components and conditions of ''evidence.'' The use of these features of TAP offer teachers a richer metalanguage for talking about science and for understanding the nature of their own discipline—and a language that we would urge to be adopted in the community, especially among those engaged in teacher training or professional development.

More importantly, our work using TAP, and our focus on the *argumentation* rather than the content of arguments themselves, has enabled the evolution of a workable framework for analysis of the quality of the process in the classroom. To date, most of those working in the field have focused on the *content* of an argument and its logical coherence. Our preference, in contrast, has been to examine the *process* of argumentation, as this is the foundation of rational thought, and to determine whether that process can be facilitated and its quality assessed.

We have also illustrated how we can apply this schema to sets of data obtained from teachers implementing argumentation in the classroom. These data sets showed evidence of positive improvement in the quality of student argumentation, but the change was not significant. This suggests that developing the skill and ability to argue effectively is a long-term process—something that comes only with recurrent opportunities to engage in argumentation throughout the curriculum rather than during the limited period of 9 months of our intervention. Our findings stand in contrast to those of Zohar and Nemet (2002), who found significant improvements after a relatively short intervention, for which we have no explanation. However, our findings are supported by the work of Zoller et al. (2000, 2002), who concluded from their work with first year college undergraduates that one semester is too short a period to develop higher order cognitive thinking and that systemic longitudinal persistence is necessary to achieve significant outcomes. The main message is that all of these studies, including our own, show that *improvement at argumentation is possible if it is explicitly addressed and taught.* Thus, it is possible for science education to make a significant contribution toward improving the quality of students' reasoning, redressing the weaknesses exposed by the work of Kuhn (1991) and Hogan and Maglienti (2001).

Finally, our data give a clear indication that supporting and developing argumentation in a scientific context is significantly more difficult than enabling argumentation in a socioscientific context. Our own view is that argumentation of quality is dependent on a body of appropriate knowledge that can form the data and warrants of an individual's arguments. In the context of socioscientific issues, students can draw on ideas and knowledge developed informally through their own life world experiences, and their ethical values. In contrast, argument in a scientific context requires very specific knowledge of the phenomenon at hand and at least a feel for the criteria for evaluating scientific evidence. Without this resource, constructing arguments of quality will be severely restricted and hampered. Thus, supporting scientific argument in the classroom requires that relevant evidence be provided to students if arguments of better quality are to be constructed and evaluated. Some will conclude that, as an argument to defend the status quo, students must acquire a knowledge of the major components of the scientific canon before they can engage in discourse activities that resemble or model those of the professional scientist. This is an argument we refute for two reasons. First, even the simplest scenarios can engage students in epistemic activities that closely model those of professional scientists. What is essential is that the process is supported by a body of relevant evidence that students can then consider and martial to support one theory or another. So, for instance, students can consider whether day and night are caused by a spinning Earth and moving Sun. Data for consideration can be that the Sun appears to move; that when you jump up you land in the same spot; that it is night time in Australia when it is daylight in Europe; that the Earth is not an exact sphere but slightly wider at the equator; that a long pendulum does not swing in the same plane all day and more. Dividing students into groups and asking them to argue the case for one view or the other, and to think how they would argue against any items of evidence that are not supportive of the theory they are defending, requires thought and develops students' critical thinking. The only legitimate moral requirement of the teacher is that they ensure that all students have some knowledge of these data—none of which are excessively demanding. The work of Keogh and Naylor (1999) on concept cartoons has shown that there are many more natural phenomena that can also be a locus of argumentation from an early age. Second, it has been our experience, and that of others (Ogborn, Kress, Martins, & McGillicuddy, 1996), that opportunities to engage in argumentation generate student engagement—the *sine qua non* of significant learning. Third, Nolen (2003) found that "in classrooms where students perceive their science teacher as interested in student understanding and independent thinking, rather than in the speedy recitation of correct answers, students are more likely to have productive and satisfying learning experiences" (p. 365).

In the next phase of our work, we have developed a set of materials to support teacher professional development in the use of ideas, evidence and argument in science, called the IDEAS project.[3] This project is rooted in the belief that a major barrier to the uptake and dissemination of such work is the lack of good examples modeling the implementation of innovative practice (Joyce, 1990). Therefore, using the teachers we have worked with in both phases of this work, we have videoed them implementing argumentation, illustrating how such lessons are organized and the key features of practice. In addition, we have developed materials for a teacher's handbook and classroom materials as well as materials that help to develop teachers' underlying theoretical understanding of the nature and function of argument in science. The latter we see as essential for developing value congruence (Harland & Kinder, 1997) that argumentation is an important aspect of science and science education.

Perhaps most significantly, however, we see our work not in isolation but as part of a growing body of work in this area (Herrenkohl & Guerra, 1995, 1998; Kelly & Crawford, 1997; Kelly et al., 1998) that has begun to explore the difficulties and dilemmas of introducing argument to science classrooms—work that attempts to offer some insight into how practice can be developed. Contemporary research has guided many educational researchers to conceive of thinking and reasoning as acts that are socially driven, language dependent, governed by context or situation, and involving a variety of tool-use and cognitive strategies. Some investigators have examined the challenges that these new ideas have about knowledge and learning for teacher education, summarizing these newer conceptions of learning as cognition as social (in that it requires interaction with other), cognition as situated (in that it is domain-specific and not easily transferable), and cognition as distributed (in that the construction of knowledge is a communal rather than individual activity), respectively. Nevertheless, a missing crucial component of this body of research is any significant evidence demonstrating that engaging in discursive problem-solving activities leads to enhanced cognition—one of the major goals of any type of education. Having established a modus operandi for argument in the classroom, and demonstrated that student skills at argumentation can be enhanced, the question we ask is: Can regular engagement in such activities over an extended period lead to enhanced cognitive development? It is this question future research in the field needs to address.

Finally, for better or for worse, the rationality of science and its commitment to evidence now permeates the discourse of contemporary life. Given science's cultural significance, exposing the nature of the arguments and epistemic thinking that lie at its heart has become a growing imperative for present-day science education. Science ‘‘for all’’ can only be justified if it offers something that is of universal value to everyone. Also, given that argumentation is a major constitutive element of science itself, and of our cultural milieu, developing some understanding of its nature and function is an essential component of the education of all young people. Engaging students in argumentation and its evaluation offers a means of transcending the dogmatic, uncritical, and unquestioning nature of so much of the traditional fare offered in science classrooms.

In short, teaching argumentation offers one means of realizing Schwab's vision that science education should be an ‘‘enquiry into enquiry’’ and perhaps, more importantly, the potential to make science education an education in critical thinking. This research offers one small contribution toward achieving such a vision.

Acknowledgments

The authors thank the many teachers who worked with us on this project. This study was supported by the UK Economic and Social Science Research Council grant number R000237915.

Notes

[1]Inevitably, due to illness and migration some of the members of each group changed.

[2]In analysing argument, our view has been that any argument that transcends a mere claim must contain an item of data. Arguments that contain only warrants without data are very difficult to construct and have rarely been observed.

[3]Funded by the Nuffield Foundation.

References

Alexopoulou, E. & Driver, R. (1997). Small group discussions in physics: Peer interaction modes in pairs and fours. Journal of Research in Science Teaching, 33, 1099–1114.

Alverman, D.E., Qian, G., & Hynd, C.E. (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. Journal of Educational Research, 88, 146–154.

Alvermann, D.E. & Hynd, C.R. (1986). Effects of prior knowledge activitation modes and text structure on nonscience majors' comprehension of physics. Journal of Educational Research, 83, 97–102.

Andrews, R. (1995). Teaching and learning argument. London: Cassell.

Austin, J.L. & Urmson, J.O. (1976). How to do things with words (2nd ed.). London: Oxford University Press.

Bachelard, G. (1940). The philosophy of no. Paris: Paris University Press.

Barnes, D. (1976). From communication to curriculum. London: Penguin.

Beck, U. (1992). Risk society: Towards a new modernity. London: Sage.

Bereiter, C. & Scardamalia, M. (1987). The psychology of written composition. Hillsdale, NJ: Lawrence Erlbaum.

Billig, M. (1996). Arguing and thinking (2nd ed.). Cambridge: Cambridge University Press.

Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, 2, 141–178.

Cohen, D. (1995). Argument is war. . .and war is hell: Philosophy, education, and metaphors for argumentation. Informal Logic, 17, 177–188.

Cowie, H. & Rudduck, J. (1990). Co-operative learning traditions and transitions. Learning together—working together. Vol. 3. London: BP Educational Service.

Dillon, J.T. (1994). Using discussion in classrooms. Buckingham, UK: Open University Press.

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

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

Erduran, S., Osborne, J.F., & Simon, S. (in press). TAPping into Argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. Science Education.

Garratt, J., Overton, T., & Threlfall, T. (1999). A question of chemistry: Creative problems for critical thinkers. Harlow, UK: Pearson.

Giddens, A. (1990). The consequences of modernity. Cambridge, UK: Polity Press.

Gilbert, J.K. & Watts, D.M. (1983). Concepts, misconceptions and alternative conceptions: Changing perspective in science education. Studies in Science Education, 10, 61–98.

Goldsworthy, A., Watson, R., & Wood-Robinson, V. (2000). Developing understanding in scientific enquiry. Hatfield, UK: Association for Science Education.

Guzetti, B.J., Synder, T.E., Glass, G.V., & Gamas, W.S. (1993). Meta-analyis of instructional interventions from reading education and science education to promote conceptual change in science. Reading Research Quarterly, 28, 116–161.

Harland, J. & Kinder, K. (1997). Teachers' continuing professional development: Framing a model of outcomes. British Journal of In-Service Education, 23, 71–84.

Herrenkohl, L., Palinscar, A., DeWater, L.S., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. The Journal of the Learning Sciences, 8, 451–493.

Herrenkohl, L.R. & Guerra, M.R. (1995). Where did you find your theory in your findings? Participant structures, scientific discourse, and student engagement in fourth grade. Paper presented at the AERA annual meeting.

Herrenkohl, L.R. & Guerra, M.R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. Cognition and Instruction, 16, 431–473.

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

Hynd, C. & Alvermann, D.E. (1986). The role of refutation text in overcoming difficulty with science concepts. Journal of Reading, 29, 440–446.

Hynd, C., McNish, M., Qian, G., Keith, M., & Lay, K. (1994). The role of instructional variables in conceptual change in high school physics topics. Journal of Research in Science Teaching, 31, 933–946.

Hynd, C.R., Alvermann, D.E., & Qian, G. (1997). Preservice elementary school teachers' conceptual changes about projectile motion: Refutation text, demonstration, affective factors, and relevance. Science Education, 81, 1–27.

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

Joyce, B. (Ed.). (1990). Changing school culture through staff development: 1990 yearbook of the Association for Supervision and Curriculum Development. Alexandria, VA: ASCD.

Kelly, G.J. & Crawford, T. (1997). An ethnographic investigation of the discourse processes of school science. Science Education, 81, 533–560.

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

Keogh, B. & Naylor, S. (1999). Concept cartoons, teaching and learning in science: An evaluation. International Journal of Science Education, 21, 431–446.

Koslowski, B. (1996). Theory and evidence: The development of scientific reasoning. Cambridge, MA: MIT Press.

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

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. Cognition and Instruction, 15, 287–315.

Kutnick, P., Blatchford, P., & Baines, E. (2002). Pupil groupings in primary school classrooms: Sites for learning and social pedagogy? British Educational Research Journal, 28, 187–206.

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

Levinson, R. & Turner, S. (2001). Valuable lessons: Engaging with the social context of science in schools. London: The Wellcome Trust.

Millar, R. & Osborne, J.F. (Eds.) (1998). Beyond 2000: Science education for the future. London: King's College London.

Monk, M. & Osborne, J. (1997). Placing the history and philosophy of science on the curriculum: A model for the development of pedagogy. Science Education, 81, 405–424.

Naylor, S. & Keogh, B. (2000). Concept cartoons in education. Sandbach, UK: Millgate House Publishers.

Nolen, S.B. (2003). Learning environment, motivation and achievement in high school science. Journal of Research in Science Teaching, 40, 4.

Norris, S.P. & Phillips, L.M. (2003). How literacy in its fundamental sense is central to scientific literacy. Science Education, 87, 224–240.

Ogborn, J. (2002). Ownership and transformation: Teachers using curriculum innovation. Physics Education, 37, 142–146.

Ogborn, J., Kress, G., Martins, I., & McGillicuddy, K. (1996). Explaining science in the classroom. Buckingham, UK: Open University Press.

Osborne, J.F. (1997). Practical alternatives. School Science Review, 78, 61–66.

Osborne, J.F. (2001). Promoting argument in the science classroom: A rhetorical perspective. Canadian Journal of Science, Mathematics and Technology Education, 1, 271–290.

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

Osborne, J.F. & Young, A.R. (1998). The biological effects of ultra-violet radiation: A model for contemporary science education. Journal of Biological Education, 33, 10–15.

Pontecorvo, C. (1987). Discussing and reasoning: The role of argument in knowledge construction. In E. De Corte, H. Lodewïjks, R. Parmentier, & P. Span (Eds.), Learning and instruction: European research in an international context (pp. 239–250). Oxford, UK: Pergamon Press.

Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. International Journal of Science Education, 19, 2, 167–182.

Ratcliffe, M. & Grace, M. (2003). Science education for citizenship. Buckingham, UK: Open University Press.

Ravetz, J. (2002). Reflections on the new tasks for science education. Unpublished evidence submitted to the House of Commons Committee for Science and Technology.

Rudduck, J. (1983). The humanities project: an introduction (revised edition). Norwich, UK: University of East Anglia School of Education Publications.

Schwarz, R. & Lederman, N.G. (2002). ‘‘It’s the nature of the beast’’: The influence of knowledge and intentions on learning and teaching the nature of science. Journal of Research in Science Teaching, 39, 205–236.

Scott, P. (1998). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. Studies in Science Education, 32, 45–80.

Simon, S., Erduran, S., & Osborne, J.F. Developing the teaching of argumentation in school science. Manuscipt submitted for publication.

Solomon, J. (1990). The discussion of social issues in the classroom. Studies in Science Education, 18, 105–126.

Solomon, J. (1991). Exploring the nature of science: Key Stage 3. Glasgow, UK: Blackie.

Solomon, J., Duveen, J., & Scott, L. (1992). Exploring the nature of science: Key Stage 4. Hatfield, UK: Association for Science Education.

Suppe, F. (1998). The structure of a scientific paper. Philosophy of Science, 65, 381–405.

Thorley, N.R. & Treagust, D.F. (1986). Conflict within dyadic interactions as a stimulant for conceptual change in physics. European Journal of Science Education, 9, 203–216.

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

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

Wells, G. (1999). Dialogic inquiry: Towards a sociocultural theory and practice of education. New York: Cambridge University Press.

White, R. & Gunstone, R. (1992). Probing understanding. London: Falmer Press.

Wray, D. & Lewis, M. (1997). Extending literacy: Children reading and writing non-fiction. London: Routledge.

Yore, L., Bisanz, G.L., & Hand, B.M. (2003). Examining the literacy component of science literacy. International Journal of Science Education, 25, 689–725.

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

Zoller, U., Ben-Chaim, D., Pentimalli, R., & Borsese, A. (2000). The disposition towards critical thinking of high school and university science students: An inter–intra Israeli–Italian study. International Journal of Science Education, 22, 571–582.

Zoller, U., Dori, Y.J., & Lubezky, A. (2002). Algorithmic, LOCS and HOCS (chemistry) exam questions: Performance and attitudes of college students. International Journal of Science Education, 24, 185–203.