TAPping into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse

SIBEL ERDURAN
School of Education, University of Delaware, Newark, DE 19716, USA

SHIRLEY SIMON
Institute of Education, University of London, London WC1H 0AL, UK

JONATHAN OSBORNE
Department of Education and Professional Studies, King's College, University of London, London SE1 9NN, UK

Received 5 November 2002; revised 10 March 2004; accepted 17 March 2004

DOI 10.1002/sce.20012
Published online 4 October 2004 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: This paper reports some methodological approaches to the analysis of argumentation discourse developed as part of the two-and-a-half year project titled "Enhancing the Quality of Argument in School Science" supported by the Economic and Social Research Council in the United Kingdom. In this project researchers collaborated with middle-school science teachers to develop models of instructional activities in an effort to make argumentation a component of instruction. We begin the paper with a brief theoretical justification for why we consider argumentation to be of significance to science education. We then contextualize the use of Toulmin's Argument Pattern in the study of argumentation discourse and provide a justification for the methodological outcomes our approach generates. We illustrate how our work refines and develops research methodologies in argumentation analysis. In particular, we present two methodological approaches to the analysis of argumentation resulting in whole-class as well as small-group student discussions. For each approach, we illustrate our coding scheme and some results as well as how our methodological approach has enabled our inquiry into the quality of argumentation in the classroom. We conclude with some implications for future research in argumentation in science education. © 2004 Wiley Periodicals, Inc. *Sci Ed* **88:**915–933, 2004

Correspondence to: Sibel Erduran; e-mail: sibel.erduran@kcl.ac.uk
Contract grant sponsor: UK Economic and Social Science Research Council.
Contract grant number: R000237915.

INTRODUCTION

Over the past few decades numerous studies have focussed on the analysis of argumentation discourse in educational contexts (e.g., Driver, Newton, & Osborne, 2000; Duschl, Ellenbogen, & Erduran, 1999; Forman, 1992; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Kelly & Takao, 2002). These studies have highlighted the importance of discourse in the acquisition of scientific knowledge (Erduran & Osborne, in press; Pontecorvo, 1987; Schwarz et al., 2003) and the development of habits of mind in science (e.g., Boulter & Gilbert, 1995; Kuhn, 1992). The implication is that argumentation is a form of discourse that needs to be appropriated by children and explicitly taught through suitable instruction, task structuring, and modeling (e.g., Mason, 1996).

Recent approaches have thus framed science learning in terms of the appropriation of community practices that provide the structure, motivation, and modes of communication required to sustain scientific discourse (Forman, 1992; Kelly & Chen, 1999; Lemke, 1990). These approaches stand in sharp contrast to the traditional views of science learning that focus on outcomes such as problem solving (Gable & Bunce, 1984), concept learning (Cros, Chastrette, & Fayol, 1987), and science-process skills (Heeren, 1990). Science learning is thus considered to involve the construction and use of tools which are instrumental in the generation of knowledge about the natural world. In this framework, argumentation is a significant tool instrumental in the growth of scientific knowledge (Kitcher, 1988) as well as a vital component of scientific discourse (Pera, 1994). Argumentation plays a central role in the building of explanations, models, and theories (Siegel, 1995) as scientists use arguments to relate the evidence they select to the claims they reach through use of warrants and backings (Toulmin, 1958).

In this paper, we report on the developments in the application of Toulmin's Argument Pattern (TAP) (Toulmin, 1958) to the analysis of argumentation discourse in science classrooms. The work cited is a component of the project titled "Enhancing the Quality of Argument in School Science" conducted between 1999 and 2002 and supported by the Economic and Social Research Council in the United Kingdom. In this project researchers collaborated with middle-school science teachers to develop models of instructional activities in an effort to make argumentation a component of instruction. The project has thus sought to elucidate the dynamics of classroom interactions that initiate and sustain argumentation discourse. Several studies have been carried out within the context of the project (e.g., Simon, Osborne, & Erduran, 2003) to determine the effectiveness of instructional interventions as well as the impact of varying subject-matter contexts on the quality of argumentation in the classroom (e.g., Osborne, Erduran, & Simon, in press).

The primary objective of this paper is to focus on the development and use of TAP as a tool for tracing the quantity and quality of argumentation in science discourse. TAP illustrates the nature of an argument in terms of claims, data, warrants, backings, and rebuttals—a framework which will be discussed in more detail in the rest of this paper. In our approach, we have adopted TAP to investigate argumentation in the whole-class discussions among teachers and students, and in small-group discussions among students. Our work extends the use of TAP in argument analysis (e.g., Pontecorvo & Girardet, 1993) by generating and applying TAP as a quantitative as well as a qualitative indicator of the teaching and learning occurring in classrooms.

We begin the paper with a brief theoretical justification for why we consider argumentation to be of significance to science education. Our choice of focussing on the scheme of argument developed by Stephen Toulmin, a philosopher, calls for a consistency of accounts from philosophy of science and cognitive studies. Hence we review some general themes on the role of argumentation in science and in cognition. We then contextualize the use of

TAP in the study of argumentation discourse and provide a justification for the methodological outcomes our approach generates. We illustrate how our work refines and develops methodologies for the analysis of argumentation in the science classroom. In particular, we present two methodological approaches to the analysis of argumentation discourse. One involves the use of TAP in quantifying arguments generated in whole-class discussions between teachers and students. This approach also provides some qualitative comparisons between arguments generated in different lessons. The second focuses on the use of TAP as an indicator of students' use of rebuttals in group work. Here, the quality of argumentation is defined in terms of the presence and nature of the rebuttals that are voiced among students. For both approaches, we illustrate our coding scheme and some results that illustrate how we have approached our inquiry into the study of argumentation in the science classroom. We conclude with some recommendations for future argumentation research in science education.

THEORETICAL BACKGROUND TO ARGUMENTATION

The philosophical and cognitive foundations of argumentation have played a central role in the justification of research in argumentation in science education (e.g., Duschl & Osborne, 2002). Contemporary perspectives in philosophy of science (e.g., Giere, 1991; Kitcher, 1988) emphasize that science is not simply the accumulation of facts about how the world is. Science involves the construction of theories that provide explanations for how the world may be. In proposing provisional explanations for the underlying causes of events, theories are open to challenge and refutation (e.g., Popper, 1959). Science often progresses through dispute, conflict, and argumentation rather than through general agreement (e.g, Kuhn, 1962; Latour & Woolgar, 1986). Thus, arguments concerning the appropriateness of experimental design, the interpretation of evidence, and the validity of knowledge claims are at the heart of science, and are central to the everyday discourse of scientists. Scientists engage in argumentation and it is through this process of argumentation within the scientific community that quality control in science is maintained (Kuhn, 1962).

Beyond coherence with current philosophies of science, there are cognitive values of argumentation in science education. From the cognitive perspective, to the extent that argument involves the public exercise of reasoning (Billig, 1987; Kuhn, 1992), lessons involving argument will require children to externalize their thinking. Such externalization requires a move from the *intra*-psychological plane, and rhetorical argument, to the *inter*-psychological and dialogic argument (Vygotsky, 1978). When children engage in such a process, and support each other in high-quality argument, the interaction between the personal and the social dimensions promotes reflexivity, appropriation, and the development of knowledge, beliefs, and values. Furthermore, to grasp the connection between evidence and claim is to understand the relationship between claims and warrants and to sharpen children's ability to think critically in a scientific context, preventing them from becoming blinded by unwarranted commitments (Quinn, 1997).

From the sociocultural perspectives on cognition, argumentation is a critical tool for science learning since it enables within learners the appropriation of community practices including scientific discourse (Kelly & Chen, 1999). If enculturation into scientific discourse is significant to science learning, then it becomes imperative to study such discourse to understand how the teaching and learning of argumentation can be traced, assessed, and supported (Duschl & Osborne, 2002). In this sense, the improvement and development of tools for capturing implementation of significant features of argumentation becomes a major concern for science education research.

ANALYZING ARGUMENTATION WITH "TOULMIN'S ARGUMENT PATTERN"

Through his well-known book titled *The Uses of Argument*, Stephen Toulmin has made a significant impact on how science educators have defined and used argument. Toulmin's definition of argument (Figure 1) has been applied as a methodological tool for the analysis of a wide range of school subjects including science (e.g., Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Zohar & Nemet, 2002), history (Pontecorvo & Girardet, 1993), and English (Mitchell, 1996). It has also been used as a heuristic for assessment of student work (e.g., Hart, 1998) as well as for supporting student learning (Andrews, 1995). For example, Mitchell (1996) has successfully adapted TAP as a heuristic to scaffold university students' writing.

In the context of science lessons, the use of TAP has mainly concentrated on the description of small-group discussions among students. For instance, Jiménez-Aleixandre, Rodríguez, and Duschl (2000) have used TAP to examine students' reasons and justifications in the context of high school genetics lessons. Duschl, Ellenbogen, and Erduran (1999), extending the work of Pontecorvo and Girardet (1993) in history topics, have examined the use of argumentative operations in student discussions on buoyancy and floatation.

Given the background into the use of TAP in science education, we will concentrate our discussion on this definition of argument in our paper although we do acknowledge the availability of other schemes of argument in the literature (e.g., Walton, 1996). Our aim in this paper is to extend the applicability of TAP in data analysis of science discourse in the classroom so the choice of TAP is central to our thesis. TAP illustrates the structure of an argument in terms of an interconnected set of a claim; data that support that claim; warrants that provide a link between the data and the claim; backings that strengthen the warrants; and finally, rebuttals which point to the circumstances under which the claim would not hold true. More specifically, in Toulmin's definition "a claim is an assertion put forward publicly for general acceptance." Grounds are "the specific facts relied on to support a given claim." Backings are "generalizations making explicit the body of experience relied on to establish the trustworthiness of the ways of arguing applied in any particular case." Rebuttals are "the extraordinary or exceptional circumstances that might undermine the force of the supporting arguments." Toulmin further considers the role of qualifiers as "phrases that show what kind of degree of reliance is to be placed on the conclusions, given the arguments available to support them."

Despite its use as a framework for defining argument, the application of TAP to the analysis of classroom-based verbal data has yielded difficulties. The main difficulty has

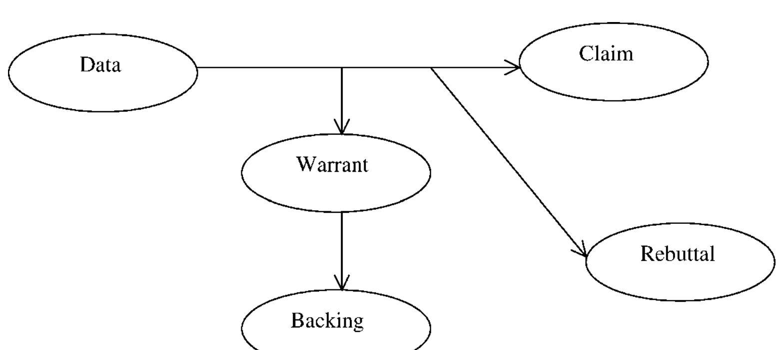


Figure 1. Toulmin's Argument Pattern (Toulmin, 1958).

been in the clarification of what counts as claim, data, warrant, and backing. Kelly, Druker, and Chen (1998) applied TAP to the analysis of student dyadic spoken discourse. This study identified the potential uses of Toulmin's method but surfaced methodological problems. The authors found that organizing student discourse into Toulmin's argument components required careful attention to the contextualized use of language. According to Kelly and his colleagues, while the Toulmin model makes distinctions among statements of data, claim, warrant, and backing, the scheme is restricted to relatively short argument structures and the argument components pose ambiguities. Statements of claims can serve as a new assertion to be proven or can be in service to another claim, thus acting as a warrant. In a subsequent study, Kelly and Chen (1999) modified Toulmin's model by drawing on the work of Latour (1987). They thus considered the epistemic status of students' claims in their writings and sorted these according to the model presented by Latour. This form of analysis allowed for the consideration of claims at multiple levels of theoretical generality and matched well with the categorical description of transactional use of language. Other researchers (e.g., Duschl, Ellenbogen, & Erduran, 1999) have preferred to use other analytical tools such as Walton's scheme on presumptive reasoning, justifying their choice on the ambiguity surrounding the key features of TAP in application to real discourse.

Let us consider the following example from our research which presents some ambiguity in the characterization of the claim, data, and warrant:

T: [Statement] A, the moon spins around, so the part of the moon that gives out light is not always facing us. Julian, A?
S1: The moon doesn't give out light.
T: Right, so that's why A is wrong. That's true. How do you know that?
S1: Because the light that comes from the moon is actually from the sun.
T: He is saying the light that we see from the moon is actually a reflection from the sun. How do we know that? Andrew?
S2: Because the moon is blocked by the . . . .

In this example, one could consider the statement "The moon spins around" as a piece of data which supports the claim "So the part of the moon that gives out light is not always facing us." One could also argue, however, that the student's choice of "A" (the statement on the card) is the main claim. In other words, "A is right" can be considered an implicit claim which is challenged by the next claim "The moon doesn't give out light." Deciding which of the statements to take as a claim (i.e. "The moon spins around" or "A is right") can thus become problematic.

One way of resolving such ambiguities is to examine the use of words such as "so" and "because." Indeed, the use of the operative word "so" which itself is implied in Toulmin's definition for reaching conclusions from data makes the first case described about highly convincing. In other words, there is little doubt that there is a claim and a justification, whatever the precise nature of this justification might be or indeed whichever statement ("the moon spins around" or "the part of the moon that gives out light is not always facing us") is taken to be the main claim. The use of the next statement "The moon doesn't give out light" as a rebuttal creates an opposition to the justification used in the primary argument. The student's further elaboration of reasoning in "Because the light that comes from the moon is actually from the sun" is an effort for a justification of the rebuttal. Viewed in this way, ambiguities about what counts as claim, data, rebuttal, and so on become less problematic. Even though all the statements above can be considered as claims in themselves, in the course of the reasoning, they can be positioned to be data or rebuttal relative to the main claim which creates an impetus for the generation of the subsequent statements.

Even though in our work the context of discourse yielded reliable coding at the level of the characterization of claims, data, warrants, backings, and rebuttals, we see this as a product of the significant time devoted to resolving disagreements. Our coding scheme has been mainly guided by a differentiation between claims, justifications, and rebuttals at the first instance, and a tighter and finer level differentiation of codes for justifications (i.e. warrants and backings) which can be complex. However, we have been able to apply the TAP scheme to the coding of a wide range of discussions from rocks to endangered species. In the next sections, we will illustrate in more detail the methodological approaches that we have adopted and the outcomes in terms of the study of teaching and learning that such approaches have enabled us to trace in the data.

METHODOLOGICAL OUTCOMES

Our adaptation of TAP for methodological purposes has created the investigation of new possibilities in our research. Even though the use of TAP in previous studies has provided valuable information regarding student reasoning and argumentation, it did not yield much insight as to how the quality of argumentation discourse might progress through sustained intervention in the classroom or indeed how TAP could be used to monitor this change. In other words, the potential of TAP to express quality of argumentation discourse in the classroom in an extended timeframe through instructional support has been a neglected component of argumentation discourse analysis. We consider it a significant weakness that the use of TAP as an indicator of improvement in argumentation quality has been understudied in science education. One of the consequences is that we have little understanding of how, for instance, TAP can be used as a quantitative as well as a qualitative indicator of argumentation over time.

In this paper, we address such concerns directly. We have developed two methodological approaches for the analysis of discourse from whole-class and small-group discussions. First, we have adapted TAP for the purposes of coding data that originate from whole-class conversations where successive implementation of lessons can be traced for their improved quality of argumentation. Here we have traced the frequency of TAP profiles from the same lessons that were implemented a year apart by the same teachers. Comparison of the results holds the potential to investigate whether or not there is an improvement in the employment of argumentation across different lessons. Our purpose here is not to report on statistically significant outcomes since our sample size was small (i.e. two lessons per teacher and no control lessons) but rather our aim is to describe a methodology that can be of use to future researchers in the quantification of arguments to test the effectiveness of interventions based on argumentation. Our analysis also provides a qualitative indication also of how teachers' specific discourse practices compare and thus how appropriate feedback can be crafted to facilitate particular teachers' implementation of argumentation. For example, the distribution of TAP profiles across the two years was very similar for each teacher but different between teachers. The tool we have developed, then, provided us with an insight into how teachers' engagement in argumentation compares and where in discourse more emphasis is needed to improve the quality of argumentation. Given the research evidence that teachers' practices improve when they are empowered by reflection and understanding on their teaching actions (e.g., Loucks-Horsley et al., 1998) such insight would help create powerful strategies for more effective implementation of traditionally unfamiliar discourse forms such as argumentation.

Furthermore, we have generated a scheme where argumentation is assessed in terms of levels which illustrate the quality of opposition or rebuttals in the student discussions in small-group format. In this approach, we have focussed on those instances where there was a clear opposition between students and assessed the nature of this opposition in terms of the strength of the rebuttals offered. We perceive the presence of a rebuttal as a significant indicator of quality of argumentation since a rebuttal and how it counters another's argument forces both participants to evaluate the validity and strength of that argument. Research evidence (e.g., Kuhn, 1970) suggests that the cognitive skill of argument is, to some extent, founded on an understanding of how to rebut an opposer's point of view. In this sense, students' ability to formulate strong rebuttals is a significant outcome for the teaching of argumentation. We have thus traced the quality of argument by focussing on the presence or absence of rebuttals. For instance, when there was opposition between students but the opposition consisted of only counter-arguments that were unrelated, we perceived this to be low-level argumentation. In other words, in these cases, there was no indication of an understanding of a rebuttal in terms of its relation to challenging the validity of the evidence and justifications offered. There was simply no reference to the components of the argument maintained by the opposition. When, however, the rebuttal was in direct reference to a piece of evidence (data, warrants, or backings) offered, thereby engaging with a presented argument, we considered this a higher level argumentation. In this methodological approach, we have thus emphasized the use of rebuttals and developed a strategy for using TAP as a measure of interactive discourse.

We will now turn to a discussion of the two approaches in the study of argumentation in whole-class discussions and small-group student discussions. Our outline will present how the mentioned methodologies hold the potential for the characterization of teaching and learning in argumentation. In so doing, we will detail how TAP can be used as a quantitative as well as a qualitative indicator of argumentation.

TAP AS MEASURE OF TEACHING AND LEARNING

Our methodological approaches in the use of TAP have thus resulted in qualitative and quantitative measure of argumentation. In Method 1, by tracing the distribution of TAPs in whole-class discussions, we have developed profiles for the lessons of each of the 12 teachers participating in our project. In this sense, this analysis has given us an indication of teaching performance at the beginning of the first and second year of the project. In Method 2, by focussing on the nature of rebuttals in small-group student discussions we have developed a method for assessing students' dialogical argumentation. Thus we were able to trace the rebuttals generated by students at the beginning and the end of the school year.

Method 1: Tracing TAP in Teacher-Mediated Argumentation

Data Sources and Coding. The primary data source was verbal conversations audiotaped in 12 classes of year 8 (ages 12–14) students. The schools were located in the greater London area, ranged from urban to suburban settings with mixed ethnic groups. Three schools were all-girls schools, one school was private, and 11 schools were state schools. The teachers were recruited through professional contacts and they were described to be effective teachers by their principals. Throughout the school year, they were trained in workshop (total of 9) where they were familiarized with the overall objectives and research design of the project. Teacher training included some recommendations for encouraging students' use of evidence to support their claims. For instance, teachers were alerted to pose questions such as "How do you know?" "What is your evidence for . . . ?" and "What reasons do you have . . . ?" A set of activities were generated (Osborne et al., 2001) to support the teaching of argumentation. Teachers were explicitly introduced to TAP and used the theoretical framework to explore applications in their classrooms. For instance, they have generated lesson materials that would structure students' writing of arguments by rephrasing claims as "my ideas are . . . " and data as "reasons for my idea are . . . ." Some of the strategies employed in the training sessions have been published as a video-based training pack subsequently financed by the Nuffield Foundation (Osborne, Erduran, & Simon, 2004).

The lessons focussed on a socioscientific topic on zoos. The topic of zoos was chosen since it is of relevance both to the curriculum and to the everyday experience of the students. Science lessons cover concepts such as extinction and preservation of species, and field trips to zoos is commonplace at the middle-school level in England. Each lesson lasted for about an hour. The main task within the lesson was to present arguments for and against the funding of a new zoo. The underlying goal of the lesson was to facilitate students' meaning making and reasoning in the context of a socioscientific issue. Audiotape recorders were wired on the teachers so as to capture their verbal contribution to the lesson as well as their interactions with the students during the group format. Audiotape recorders were also placed at the table of two groups of students to capture the group talk as a subset of the class talk. Each lesson had three sections. At the onset, the teacher distributed a letter from a fictitious funding agency contacting the students and outlining the task. Initially there was a whole-class discussion about the pros and cons of zoos. Then the students were put into groups and asked to come to some consensus about 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. As homework, the students were typically asked to write a letter or compose a poster that would communicate their arguments.

Audiotapes were transcribed and analyzed for each teacher. In particular, we were interested in comparing the nature of arguments generated in the classroom across the two years by the teachers and the differences that might be across the teachers. Such comparisons provide one a means for determining the development of the teachers. In the case of the following example,

"Zoos are horrible, I am totally against zoos"

our focus would be on the substantive claim. In this case, the difficulty lies in the fact that both can be considered to be claims, i.e.

"'Zoos are horrible' and 'I am totally against zoos'"

The question for the analysis then becomes 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 is reliant on listening to the tape and hearing the force of the various statements here. Part of this might be substantiated by Austin and Urmson's (1976) distinction between locutionary statements—those which have an explicit meaning—and perlocutionary statements—those which have an implicit meaning. And the perlocutionary force with which these statements are distinguished is an aid to resolving which is intended as the substantive claim.

Here our reading is that the emphasis lies on the second part of the statement because the task context demands a reference to a particular position (for or against zoos) and that this is therefore the substantive claim. In choosing to use TAP in this manner, we have developed a good reliability (more than 80%) between the coders.

As an example, we will consider the following case between the student and the teacher.

S: I've got a con. If the animals are always walking about in the same places they might get angry and be dangerous.

T: Right, this is an anti, is it? So, being caged may alter their behavior.

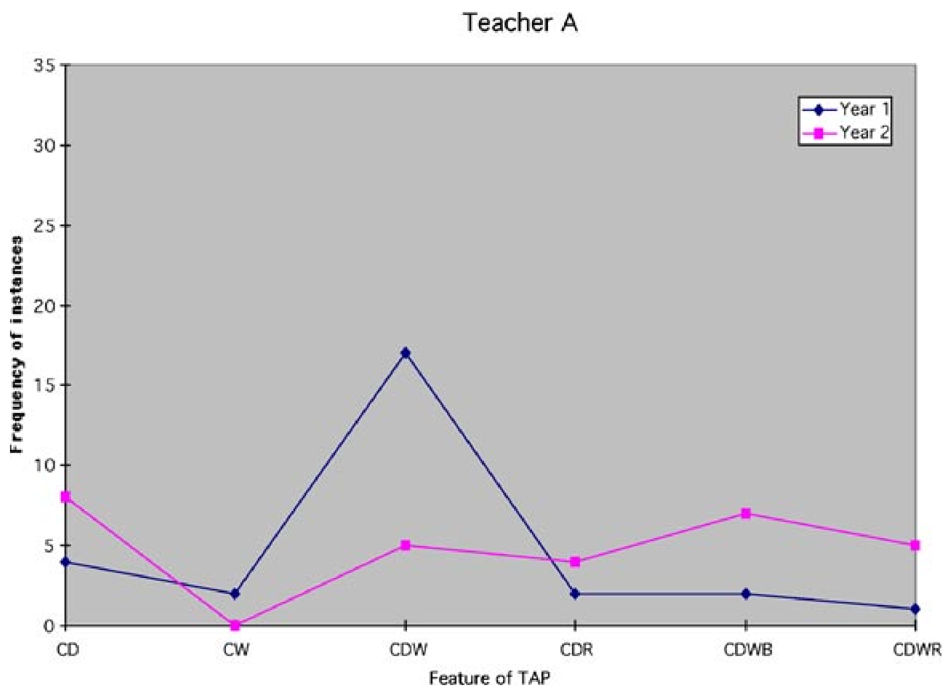
The position represented by the student is "against zoos" or "con" which is the central claim: "I've got a con." The student further adds onto this claim by saying that "if the animals are always walking about in the same places, they might get angry and be dangerous." We consider this elaboration as data to support his claim. The teacher subsequently interprets and justifies the choice for data by saying that "being caged may alter their behavior." We regard the teacher's contribution as the warrant to the argument being constructed. Such a coconstruction of arguments between students and teachers was typical in all the transcripts we have studied using the wired tapes.

Whatever the appeal that was subsequently made (e.g., "Zoos are horrible because they treat animals badly and when treated badly animals suffer psychologically") was structured enough to follow a data–warrant–backing sequence (e.g., claim: I am against zoos; data: "Zoos are horrible"; warrant: "Because they treat animals badly"; and backing: "When treated badly animals suffer psychologically.") The elaboration of this sequence was typically structured through the use of writing frames which asked students to state their position (claim), their reasons for their position (data), the ideas that support their reasons (warrant), and further information that supports these latter ideas (backings). Teachers also encouraged opposition through reflective talk in the classroom about opposing points of view. For instance, one of the strategies that the teachers were asked to trial was to select two students who had different points of view and position them to argue against each other. The teacher here made use of questions such as "How would you argue against that?" "What evidence would you provide to show him that his idea is wrong?" Hence the emergence of rebuttals was also built explicitly into this lesson through reference to the particular positions that students took already.

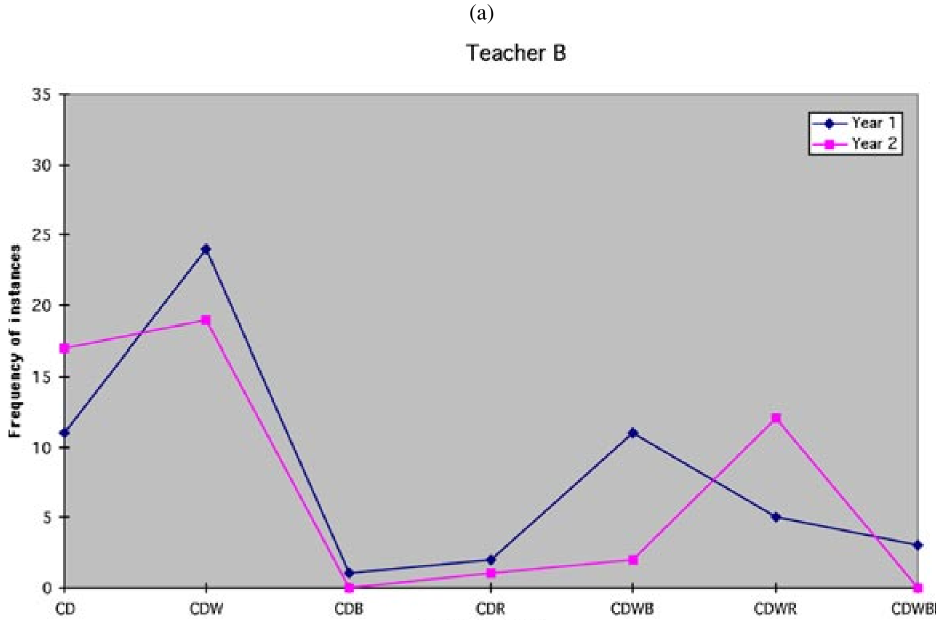
Results. Typically, the TAPs were generated between students and teachers whereby for instance, a student would provide a claim or data and a teacher would provide a warrant for this claim–data pair. Once the transcripts were coded, the trends in the distribution of TAP in each lesson were traced in the following fashion. First, we aimed at identifying the "argument space" in each lesson. By argument space, we mean the nature and frequency of TAPs occurring in a particular lesson. Second, our intention for tracing arguments spaces of lessons was twofold: (a) we wanted to investigate how the particular aspect of TAP was emphasized in each lesson; and (b) given the trends in the frequency and permutation of TAP we wanted to detail the pedagogical strategies that could enhance the teaching of argumentation. For instance, we have examined how the TAP profiles related to particular aspects of each teacher's talk which we have reported elsewhere (Simon, Osborne, & Erduran, 2003).

Each teacher implemented the same activity one year apart with comparable students. In other words, the students in each school across the two years came from the same neighborhood with similar ethnic, linguistic, and racial backgrounds. The lessons were similar in structure, i.e. there was an introduction, group discussions, group presentations, and finally assignment of homework in either case for both years. Some examples of coded data are summarized in the charts in Figure 2. The x-axis indicates the features of TAP that were used in different combinations. For example, CD indicates those instances where a claim was coupled with data and no other feature of TAP. CDWB indicates that there was a claim, data, warrant, and backing. The y-axis illustrates the frequency of instances that such permutations of TAP occurred within the transcript. In other words, we counted the number of times that any single argument was formulated in terms of whichever combination of TAP features.

The figures suggest several trends. First, there is argumentation in the classrooms of these three teachers across both years. This was the case for all of the 12 teachers involved in the project. In the figures, we see specific examples of to what extent each teacher's class is involved in the construction of the particular aspect of TAP. We can trace the nature of different permutations of TAP in either teacher's implementation of the lesson. Second, each teacher's classroom lays out an argument space in the same way across the two years. In other words, the trends across the use of different permutations of TAP are similar across



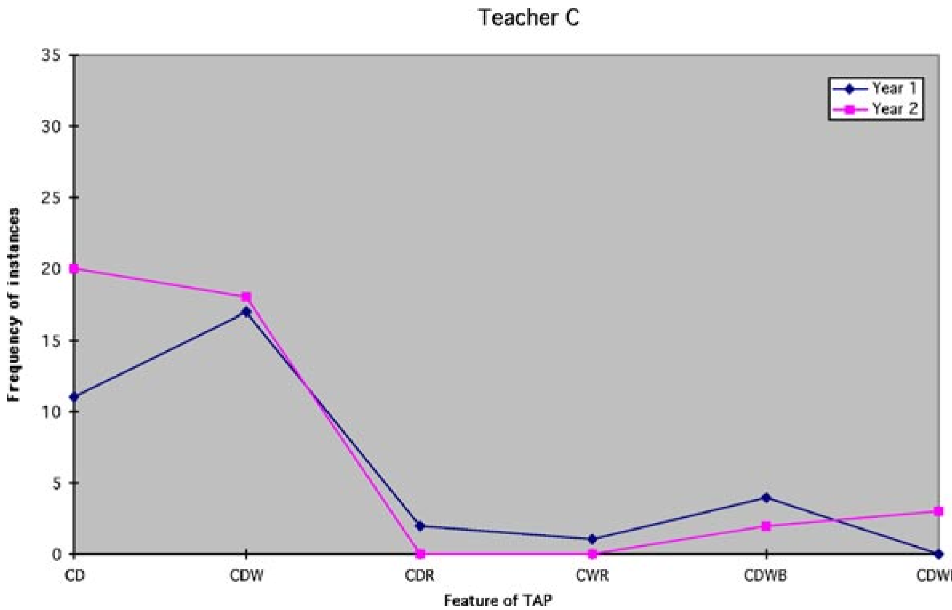
(a)



(b)

Figure 2. Distribution of TAP for (a) Teacher A, (b) Teacher B, and (c) Teacher C.

Continued



(c)

Figure 2. *Continued.*

two years. Since the transcripts were coded one year apart and randomly across the two years, there is little likelihood that the similarities in patterns indicate coder bias. Third, the figures suggest that there is no common pattern across all teachers and that there are no universals to the nature of argument space generated in the project classrooms: the use of argument seems to be teacher dependent. There seems to be a teacher-specific effect to the profile of argumentation discourse—even though the students were different across two years and much of the TAP contributions come from students.

Results from TAP coding of each transcript were further summarized in the following fashion. First, the TAPs were grouped in terms of the occurrence of double, triple, quadruple, and quintuple combinations. We called these permutations of TAP features "clusters." For example, the "claim–data–warrant" and "claim–data–rebuttal" were grouped as an instance of cluster 3. It is implicit in our grouping that with increasing number for a cluster, the argument becomes more complex in nature. In other words, we are assuming that a "claim–data" argument is a less sophisticated form of an argument than a "claim–data–warrant" argument where there is an added feature of justification (in terms of a warrant) in the latter scenario. Furthermore, for our coding purposes, we concentrated on identifying arguments in terms of the quantity of TAP features in arguments, not qualitative differences across different permutations of TAP. That is to say, by collapsing different arguments into clusters, we are not differentiating between arguments that might have a different qualitative composition despite a quantitative equivalence in terms of TAP, i.e. "claim–data–warrant–rebuttal" and "claim–data–warrant–backing," both instances of cluster 4, are grouped together since each has four features of TAP even though qualitatively there is a difference between the arguments in terms of the presence or absence of rebuttals and backings.

We traced the percent frequency distribution of the clusters for each teacher for both years of the project. Teachers' individual differences in emphasis on different kinds of clusters might illustrate the tendencies in their understanding of what counts as argument. Interviews we have conducted with teachers will provide direct evidence regarding the beliefs teachers hold on argumentation. We intend to compare and correlate data on TAP profiles with teachers' understandings of argumentation explicated through interviews. For instance, we will pursue questions such as "How do teachers' understanding of argument relate to the TAP profiles from classroom discourse?" One hypothesis would be to expect shifts to higher order clusters as teachers' understanding of argumentation gets more sophisticated. In other words, there may be a correlation between teachers' understanding of argumentation and how they are enabling the manifestation of more complex arguments in their classrooms.

The cluster analysis showed how the discourse of the classroom is dominated by arguments that contain fewer elements of TAP and are less elaborated. Overall, when the data were collapsed across all teachers, there was a significant ($p < 0.01$) difference between year 1 and year 2 with more elaborated arguments being used in the second implementation of the zoo lesson. In other words, clusters that included two and three components (e.g., CD and CDW) occurred at a less frequency than clusters that included four and five components (e.g., CDWB and CDWBR). When we looked at differences at an individual level, analysis showed that this change is a result of the changes made by 8 of the 12 teachers and that for 4 teachers there was no significant difference. However, since our sample size was small and we did not have control lessons for each teacher, we could not implement a rigorous experimental design. Although our interpretation of the results is limited, we do believe that the methodological approach we have presented will be of use to other researchers who might be interested in conducting large-scale studies on argumentation. The methodology we have presented enables both qualitative and quantitative investigation of argumentation in classroom discourse—qualitative in terms of particular patterns in the distribution of TAP clusters and quantitative in terms of the statistical comparison of TAP cluster frequencies.

Method 2: Tracing Rebuttals in Student Group Discussions

Previous research on argumentation has consistently found the application of Toulmin's scheme problematic, as his criteria do not assist the ready resolution of data from warrants, nor warrants from backings resulting in poor reliability (Duschl, Ellenbogen, & Erduran, 1999; Kelly & Takao, 2002). In our work also we have encountered difficulties with the distinction between data and warrants, or warrants and backings although there was little problem in distinguishing claims or rebuttals. In a range of lesson contexts including a wide variety of science topics, where the structure to argument in terms of TAP was not as rigidly specified as in the zoo lesson, we tried to transcend this problem of ambiguity in TAP by concentrating on the quality of rebuttals only. Our schema for student group argumentation therefore avoids the necessity to resolve the problems which arise from the use of a generalized analytical framework in a context where meaning may be indeterminate.

In establishing this framework, we have drawn two major distinctions. The first is whether an argument consists of any reasons, i.e. data, warrants, or backing, to substantiate its claim, given transcending mere opinion and developing rational thought is reliant on the ability to justify and defend one's beliefs. The second is whether an argument consists of a rebuttal. Conversation with rebuttals are, however, of better quality than those without given that individuals who engage in talk without rebuttals remain epistemically unchallenged. The reasons for their belief are not questioned and are simply opposed by a counter-claim that may be more or less persuasive but is not a substantive challenge to the original claim. At its worst, such arguments are reducible simply to the enunciation of contrasting belief systems. For instance, given that beliefs rely on justifications using data and warrants, a confrontation between a creationist and a Darwinist without any attempt to rebut the data or the warrants of the other would have no potential to change the ideas and thinking of either. Only arguments which rebut these components of argument can ever undermine the belief of another. Oppositional episodes without rebuttals, therefore, have the potential to continue forever with no change of mind or evaluation of the quality of the substance of an argument. Thus, arguments with rebuttals, we believe, are an essential element of better quality arguments and demonstrate a higher level capability with argumentation. Furthermore, rebuttals can also be considered as a measure of conversational engagement. In other words, since one of the goals in promoting argumentation in science lessons is to engage learners in dialogical conversation where they can not only substantiate their claims but also refute others' with evidence, the presence of rebuttals in conversation can act as an indicator of sustained engagement in argumentation discourse.

Data Sources and Coding. Of the 12 teachers who participated in the first year of the project, 6 were selected to continue in the second year. The reason for the reduction in the number of teachers was financial: We had limited funds to collect data in the second phase of the project. The 6 teachers were selected on the basis of their effectiveness in promoting argumentation in their classrooms in the first year of the project. In each class of the 6 teachers, two groups of three to four pupils were identified by the teacher and their discussions were audiotaped and transcribed. The main criterion for the selection of the students was their regularity in attending school. The transcripts were then searched to identify episodes of opposition and dialogical argument. Opposition took many different forms and many arguments were coconstructed where students provided data or warrants for others' claims.

The data were obtained from several lessons: (a) a lesson on zoos conducted at the beginning of the school year during the second year of the project; (b) two science lessons, one at the beginning and the other at the end of the second year; and (c) a lesson on leisure centers, a socioscientific issue similar to the zoo scenario. In the leisure center activity, the task was to argue for or against the funding of a new leisure center to be built in an area rich in wildlife. The context created similar arguments to the zoo lesson in terms of the preservation of wildlife and promoting education about nature. The science lessons ranged in topic from teacher to teacher since each teacher adapted the argumentation work into their school's curriculum. The topics included energy, light, acids and bases, and electricity. For examples of lessons with science topics please refer to the Resource Pack generated as part of the IDEAs Project (Osborne, Erduran, & Simon, 2004).

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 where students were clearly against each other were traced. Typically these instances were identified through the use of words such as "but," "I disagree with you," "I don't think so," and so on. 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, who was elaborating on what idea or reinforcing or repeating an idea. In this paper we will only report on the nature and frequency of the opposition in terms of the quality of rebuttals offered. In subsequent studies, we intend to report more extensively the interactional analyses we have carried out using the group discussion data. Here we will briefly mention the interaction analysis which holds much potential for understanding what kinds of group dynamics might facilitate better argumentation among students. The main processes identified in such episodes were opposing claims by other, elaboration of an earlier idea, reinforcement of a claim with additional data, warrants, advancing claims, or adding qualifications. Again, such analysis helps to identify the features of the interaction and the nature of the engagement between the students.

TABLE 1
Analytical Framework Used for Assessing the Quality of Argumentation

Level 1	Level 1 argumentation consists of arguments that are a simple claim versus a counter-claim or a claim versus a claim.
Level 2	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backings but do not contain any rebuttals.
Level 3	Level 3 argumentation has arguments with a series of claims or counter-claims 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 counter-claims.
Level 5	Level 5 argumentation displays an extended argument with more than one rebuttal.

Each oppositional episode was analyzed using TAP to identify the principal components of an argument contributed by the individuals in the group. All episodes were read independently by two coders who then met to compare their analysis and resolve differences in interpretation. 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 is how to define their quality. What, for instance, makes one better than another? To answer this question, we have developed a framework on quality in terms of five levels of argumentation summarized in Table 1.

The following set of examples are provided to illustrate how our analysis has been applied to the data.

Episodes Without Rebuttals. The first example in Figure 3 is a short simple disagreement. This episode is simply a claim for zoos—"right for" followed by a counter-claim "we are not for it" repeated by "I am not for it," making it an example of Level 1 argumentation because the claim is unsupported by any data or warrants, and there are no rebuttals. Instead, there is simply a counter-claim and as such, there is no potential for the justification of belief to be examined and, hence, no possibility or resolution.

The second example in Figure 4 is much more complex as it involves one student providing a relatively sophisticated argument which does not appear to be understood by his opposer.

Here, what we have is a claim that "professional zoos would not hurt animals," which is countered by the claim that "animals in zoos might be scared" (claim) as "they would see other sedated animals being dragged off" (data). Despite some embedded complexity, as an example of arguing we would contend that it is essentially weak as 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 a Level 2 argumentation.

Episodes with Rebuttals. The episode in Figure 5 begins with the implicit claim that zoos are beneficial. The data for this argument is that "some animals wouldn't be able to

S1:	Right 'for'.
S2:	We are *not* for it
T:	First, write in then, then write things around it
S2:	I am *not* 'for' it.

Figure 3. Example of Level 1 argumentation.

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.

Figure 4. Example of Level 2 argumentation.

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 is that otherwise "they will be at risk from predators." This second claim is weakly rebutted with a negation which 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.

As an example for a Level 4 argument, consider the episode in Figure 6. This example, referred to in the introduction of this paper, was taken from a scientific context where pupils have been given alternative theories to explain the phases of the moons that are on numbered card, A, B, C, D, which are referred to in the dialogue.

Here, the first pupil advances the claim that it is explanation A appealing to a datum that "the moon does not give out light." 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 which is unfinished.

Results. The chart in Figure 7 shows the distribution of the levels of arguments obtained from 43 discussion groups in 23 lessons. A complete set of data (six teachers, four lessons per teacher) is not available due to technical difficulties in the data collection. At each level of argumentation, data from the beginning of the school year (zoo lesson and a science

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.........

S3: Extinction.

S1: The animal needs a place to live because they would be at risk from other predators.

S2: What are you putting?

S1: A place to live, or they would be at risk from other predators.

S1: They might not have enough food to eat.

S2: But I mean, that's nature, one has to....

S1: But we are for it

Figure 5. Example of Level 3 argumentation.

T	.... A, the moon spins around, so the part of the moon that gives out light is not always facing us. Jamal, A?
S1	The moon doesn't give out light.
T	Right, so that's why A is wrong. That's true. How do you know that?
S1	Because the light that comes from the moon is actually from the sun.
T	He is saying the light that we see from the moon is actually a reflection from the sun. How do we know that? Mark?
S2	Because the moon is blocked by the…..

Figure 6. Example of Level 4 argumentation.

lesson) are represented by "pre" while the data from the end of the school year (leisure center and a science lesson) are represented by "post." This chart shows that the largest number of arguments emerging from the data both at the beginning and at the end of the year was at Level 2 (38% and 30%, respectively). Encouragingly though, whereas at the beginning of the year only 40% of pupil arguments were at Level 3 or above at the beginning of the year, by the end of the year, the corresponding figure was 55%. Whilst we are not able to make any judgments based on statistical significance, the trends suggest a positive development in the quality of argument. Moreover, the number of Level 1 arguments has reduced from 22% to 15%. This finding is particularly encouraging as it suggests that only a small minority of arguments developed by pupils did not attempt to offer a rationale or some grounds for their claims, and that the intervention has led to a diminishment in the number of such arguments.

This method of analysis permits a number of comparisons of the performance of the groups which we have undertaken and are reporting fully in Osborne, Erduran, and Simon (in press). Our intention here is to report on the methodological approach we have developed that enables the study of argumentation in small-group student discussions. It is important to note at this point some of the assumptions on which our data analysis has been based. In our analysis of group discussions, we have assumed a context for argumentation

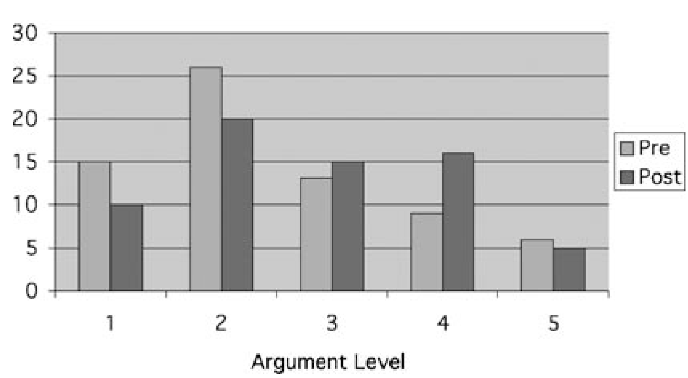


Figure 7. Chart showing numbers of each level of argumentation at the beginning and end of year lessons ($n = 43$).

where there is interactive conversation with alternative points of view which by its nature invites rebuttals. Dialogical argumentation was a feature that was explicitly promoted by our project teachers, for instance through activities that used tasks that included competing theories and alternative explanations for particular phenomena. If, however, the argumentation context is not dialogical we believe that rebuttals are still crucial in the study or argumentation. The extent to which even a single argument can anticipate potential opposition is something we perceive as a higher order skill than if such anticipation did not exist. Indeed, Toulmin himself has argued for this position, that a good argument even as presented by a single individual in a rationalized way would have considered potential circumstances under which the main claim might not hold true (Toulmin, 1958). For instance, in illustrating the structure of an argument, he gives the example of "Harry is a British subject" as a main claim and "Unless he was naturalized American" as a potential rebuttal. The rebuttal directly addresses the evidence presented as data ("He was born in Bermuda") and warrant ("Everyone born in Bermuda becomes a British citizen") to refute the original claim.

CONCLUSIONS AND IMPLICATIONS

In this paper we have outlined two methodological approaches that extend the use of Toulmin's model for tracing argumentation discourse in science classrooms. While we acknowledge the problems that TAP presents, we believe that our schemes improve the use of TAP in a significant way. Our methodological tools extend the measurement of the qualitative and quantitative outcomes of teaching and learning argumentation for several reasons.

First, previous studies have concentrated on the application of TAP at the level of particular segments of classroom discourse, whereas this study illustrates how coding of whole-classroom conversations can yield argument profiles which can act as indicators of improved performance across implementation of lessons. In other words, our scheme moves the use of TAP to a level where argumentation in entire lessons can be traced and examined in detail. Second, we have illustrated TAP's potential to illustrate the distribution of arguments in discourse. In other words, we have exploited the potential of TAP for the quantitative measure of not only TAP but also overall distribution of argumentative talk. Future studies could make sure of this approach to map different phases of lessons (i.e. introduction, group work, whole-class discussions) to examine if and how arguments might dominate certain parts of lessons and why. Third, our scheme shows how teachers' experiences will need to be biased toward better implementation of argumentation in the classroom, i.e. where more work is needed to move the conversation to be more inclusive of TAPs that include backings and rebuttals.

Previous efforts in the use of Toulmin's Argument Pattern (TAP) in the classroom (e.g., Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Pontecorvo, 1987) have presented a lack of resolution for tracing teachers' changing practices and children's enhanced argumentation given the TAP framework. In our work, we have extended the use of TAP to serve the purpose of judging enhanced quality of argumentation. In other words, we have selected the use of TAP as an indicator of changing argumentation. For instance, if most of the warrants to student arguments are provided by the teachers initially, it could be expected that as teachers' skills improve, their students will be more encouraged to justify their claims and provide their own warrants. In subsequent studies then, we intend to report on how teachers' argument profiles map to other indicators of effectiveness in teaching argumentation.

In summary, we have used TAP as an indicator of quality and quantity of argumentation in classroom discourse. In so doing we have developed a means for tracing improvements in argumentation over time. Furthermore, the methods used enable scaling up from individual to collective arguments and have implications for collective reasoning behaviors. Coconstruction of arguments between teachers and students provides an illustration of collaborative cognition where meaning-making processes in discourse analysis cannot be assumed independent of the nature of the local contexts, a position consistent with contemporary perspectives in situated cognition (e.g., Lave & Wenger, 1991). Discourse analysis assumes that the resources and strategies (e.g., grammar, rhetorical formations, cultural narratives) used in producing discourse events and texts are characteristics of the community in question. However, by illustrating the nature of teaching and learning interactions, discourse analysis can provide the tools for understanding how science education can be improved in general. Overall preliminary results indicate that collective reasoning is influenced strongly by the nature of teaching, as suggested by the outcomes of Method 1, a finding that can be unpacked further in future studies to examine how engagement in argumentation discourse can improve science teaching and learning.

We would like to acknowledge the efforts and ideas of the teachers as well as the students who have been involved in the project: James Bunn (Braintree), Paul Drayton (Haggerston), Mona Evan (Rooks Heath), Sue Frearson (St. Albans), Jim Henderson (Camden School for Girls), Peter Kauffman (Coopers Co. and Coborn School), Martina Lecky (Greycoat School), Alex Manning (Hornsey Girls), Sue Parkyn (Hampstead), John Spokes (Whitmore), and Mike Terry (Copthall).

REFERENCES

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

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

Billig, M. (1987). Arguing and thinking: A rhetorical approach to social psychology. Cambridge: Cambridge University Press.

Boulter, C. J., & Gilbert, J. K. (1995). Argument and science education. In P. J. M. Costello & S. Mitchell (Eds.), Competing and consensual voices: The theory and practice of argumentation. Clevedon: Multilingual Matters.

Cros, D., Chastrette, M., & Fayol, M. (1987). Conceptions of second year university students of some fundamental notions of chemistry. International Journal of Science Education, 10, 331–336.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of argumentation in classrooms. Science Education, 84(3), 287–312.

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

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

Erduran, S., & Osborne, J. (in press). Developing arguments. In S. Alsop, L. Bencze, & E. Pedretti (Eds.), Analysing exemplary science teaching: Theoretical lenses and a spectrum of possibilities for practice. Philadelphia: Open University Press.

Forman, E. A. (1992). Discourse, intersubjectivity and the development of peer collaboration: A Vygotskian approach. In L. T. Winegar & J. Valsiner (Eds.), Children's development within social contexts: Metatheoretical, theoretical and methodological issues (Vol. 1, pp. 143–159). Hillsdale, NJ: Erlbaum.

Gable, D., & Bunce, D. (1984). Research on problem solving in chemistry. In D. Gabel (Ed.), Handbook of research on ecience teaching and learning (pp. 301–326). New York: Macmillan.

Giere, R. (1991). Understanding scientific reasoning (3rd ed.). Fort Worth, TX: Holt, Rinehart, and Winston.

Hart, C. (1998). Doing a literature review: Releasing the social science research imagination. London: Sage.

Heeren, J. K. (1990). Teaching chemistry by the Socratic Method. Journal of Chemical Education, 67(4), 330–331.

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

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

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

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

Kitcher, P. (1988). The child as parent of the scientist. Mind and Language, 3(3), 215–228.

Kuhn, D. (1992). Thinking as argument. Harvard Educational Review, 62, 155–178.

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

Latour, B. (1987). Science in action: Scientists and the state. Princeton, NJ: Princeton University Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts (2nd ed.). Princeton, NJ: Princeton University Press.

Lave, J., & Wenger, E. (1991). Situated learning, legitimate peripheral participation. New York: Cambridge University Press.

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

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

Mason, L. (1996). An analysis of children's construction of new knowledge through their use of reasoning and arguing in classroom discussions. Qualitative Studies in Education, 9(4), 411–433.

Mitchell, S. (1996). Improving the quality of argument in higher education interim report. London: Middlesex University, School of Education.

Osborne, J. F., Erduran, S., & Simon, S. (2004). Ideas, evidence and argument in science. In-service Training Pack, Resource Pack and Video. London: Nuffield Foundation.

Osborne, J. F., Erduran, S., & Simon, S. (in press). Enhancing the quality of argument in school science. Journal of Research in Science Teaching.

Osborne, J. F., Erduran, S., Simon, S., & Monk, M. (2001). Enhancing the quality of argument in school science. School Science Review, 82(301), 63–70.

Pera, M. (1994). The discourses of science. Chicago: University of Chicago Press.

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: Pergamon.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. Cognition and Instruction, 11(3/4), 365–395.

Popper, K. (1959). The logic of scientific discovery. London. Hutchinson.

Quinn, V. (1997). Critical thinking in young minds. London: David Fulton.

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

Siegel, H. (1995). Why should educators care about argumentation? Informal Logic, 17(2), 159–176.

Simon, S., Osborne, J., & Erduran, S. (2003). Systemic teacher development to enhance the use of argumentation in school science activities. In J. Wallace & J. Loughran (Eds.), Leadership and professional development in science education: New possibilities for enhancing teacher learning (pp. 198–217). London & New York: RoutledgeFalmer.

Stein, N., & Miller, C. A. (1991). I win-you lose: The development of argumentative thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), Informal reasoning and education, Hillsdale, NJ: Erlbaum.

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

Vygotsky, L. (1978). Mind in society. London: Harvard University Press.

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

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