Science & Education **12:** 341–363, 2003.

Assessment of Evidence in University Students' Scientific Writing

ALLISON Y. TAKAO and GREGORY J. KELLY
Department of Education, University of California, Santa Barbara, CA 93106, USA (E-mails: atakao@education.ucsb.edu; gkelly@education.ucsb.edu)

Abstract. The purpose of the study presented in this paper is to examine uses of evidence in university students' writing of scientific argument. The study was conducted in an introductory level oceanography course in a large public university. In this course students draw on large-scale geological data sets provided by an interactive CD-ROM, 'Our Dynamic Planet' (Prothero 1995), to write a scientific technical paper. We examined the uses of evidence in two ways. First, we conducted interviews with undergraduate oceanography students $n = 9$, non-oceanography undergraduate students $n = 8$, and course instructors $n = 4$ to assess the differences in interpretation of uses of scientific evidence for high ranked and low ranked student papers from a previous academic year. Second, we applied our proposed argumentation analysis to student writing to examine the formulation of evidence in these high and low ranked papers. We draw on these findings to discuss ways of teaching students the construction of argument in scientific writing.

In this paper we examine the argumentation structure of student scientific writing in a writing-intensive university physical oceanography course. Through this analysis we identify ways of making the writing genre of geological sciences more readily accessible to students. Uses of evidence in writing are examined in two ways. First, we considered how populations with varied knowledge and experience in geological sciences assess the merits of student written argument. This was accomplished through a set of research interviews with course instructors, university oceanography students, and non-oceanography university students. In these interviews, the participants were asked to review one high and one low ranked paper from a previous academic year and identify and describe their overall opinion of the two papers, as well as specific issues such as authors' use of evidence, authors' use of figures, and conclusions made by the student authors. Second, we propose and apply an argumentation model to make visible the differences between the student papers. The argumentation model makes visible the knowledge of the disciplinary practices embedded in written knowledge and suggests implications for instruction.

Research on writing in professional communities has identified the importance of social practices and community norms in defining relevant rhetorical features required for specific writing purposes within disciplinary contexts (Schwegler &

Shamoon 1991). One key feature in scientific writing is a community's assessment and determination of what constitutes evidence in a given historical moment (Bazerman 1988). The importance of evidence in science and the recognition of the defining role of disciplinary communities has led educational research examining writing to learn issues in science education to bring epistemological issues to the foreground (Keys 1999; Prain & Hand 1999). Furthermore, evidence formation has been seen as a key dimension to scientific reason (Duschl 1990) and central to student learning of the practices of an objective epistemic community (Grandy 1997). This study builds from an ongoing ethnographic study of discipline specific scientific writing in a university oceanography course (Kelly et al. 2000; Kelly & Takao 2002) by examining the epistemological issues associated with formulating evidence in writing.

1. Rhetoric, Argument, and Evidence in Science and Education

The theoretical framework for this study builds on research of writing to learn science, argumentation in science, and the rhetoric of science more generally. Our review of recent literature in science education indicates a newly burgeoning interest in research on learning to write in science genres (Keys, 1999; 2000). To date, few research studies have focused on discipline-specific scientific writing, fewer still on geological science. Our study aims to contribute to this relatively under-researched field (Bezzi, 1999).

In a review of the research in the field of 'writing to learn science', Keys (1999) called for the use of scientific genres in instruction and the examination of classroom activities that encourage 'integrated inquiry and writing' (p. 128). Writing of this sort suggests a tie between the shaping of written knowledge and epistemological issues related to the use of evidence. Prain and Hand (1999) investigated students' perceptions of science, science writing, and learning and found that students were not able to explain how knowledge claims were established nor how 'writing could act as an epistemological tool' (p. 160). While the relationship of writing and knowledge production is currently under-developed, Prain and Hand suggested the need for research examining how students understand what counts as knowledge in science lessons, particularly as related to their own writing. One approach to developing a more epistemological orientation to writing in science in to promote the uses of argument among student writers. We therefore considered both research on argumentation as well as research on uses of persuasion in professional scientific writing.

The second research field informing our work is derived from studies of argumentation. These studies have analyzed students' and teachers' arguments to consider issues of student reasoning, engagement in scientific practices, and development of conceptual and epistemic understandings (Jimenez-Aleixandre et al. 2000; Kelly et al. 1998; Kuhn 1992). Studies have examined evidentiary authority in teacher-student discourse (Russell 1983; Carlsen 1997), ways students reason about socio-scientific issues (Patronis et al. 1999), and the appropriation of scientific discourse in students' small group conversations (Richmond & Striley, 1996).

The process of building evidentiary-based arguments moves students beyond memorization and recitation towards a more comprehensive understanding of scientific concepts. In the process of building scientific arguments students must separate important data from distracting noise when considering what evidence is most relevant to use in their argument. Furthermore, constructing scientific arguments requires knowledge and use of scientific theory (Duschl 1990). Spoken and written argumentation are two modes that allow students to externally engage (internal engagement occurs in the mind of individuals) in argumentation practices (Kuhn 1993). Studies of classroom discourse indicate that talk enables students to further develop their understanding of scientific ideas and under certain conditions allows students to learn the language of science (Driver et al. 2000; Lemke 1990). Through these two modes, spoken and written argumentation, students are provided with opportunities to engage in scientific practices by interpreting evidence to construct and reconstruct their own knowledge of the subject matter. Students must be able to understand, interpret, and make sense of their data in order to build strong arguments. This encourages the comprehension of what claims can be made known from their evidence, rather than simply repeating the 'facts'. Nevertheless, Newton et al. (1999) have argued that although argumentation is key in scientific processes, it still receives minor attention in school science and that students are given few opportunities to engage in evidence-based argumentation practices.

Although the writing to learn field has identified the writing of substantive arguments as one of the scientific genres and advocate for students to engage in such (Keys 1999), few studies have applied argumentation analysis to examine students' use of evidence in writing. Our study is situated in a writing intensive course and is therefore concerned with examining written argumentation rather than spoken argumentation. The centrality of writing in establishing scientific knowledge has been foregrounded by science studies (e.g., Bazerman 1988), and in particular, the rhetoric of science (Gross 1990; Perelli 1989), a field to which we now turn.

The third part of our theoretical framework draws from studies in the rhetoric of science[1] – 'the study of how scientists argue in the making of knowledge' (Harris 1997, p. xii). The production of written texts has played a central role in scientific communities and analysis of the history of the cultural practices associated with the production of such texts has identified how the uses and purposes of written knowledge have changed with changing mores in scientific communities (Atkinson 1999; Bazerman 1988; Swales 1990). Rhetorical studies of science view knowledge as actively constructed by scientists working individually or collectively on problems and being held accountable to public standards. Such public standards require that authors of knowledge claims articulate their reasoning, marshal appropriate evidence given a particular topic within a discipline, and recognize the limits of their arguments. Claims to knowledge are often formally made through written language (Bazerman 1988). Given the local aspects of persuasion as well as the constraints posed by the norms of the relevant audience (Gieryn 1999), knowledge formation can be perceived as more than an individual function; rather it is part of a communal engagement with the material world (Goodwin 1995).

These studies of the rhetoric of science have specific implications for science students. Student writers of scientific arguments need to develop facility with the key concepts, theoretical commitments, and typical uses of empirical evidence within specific disciplinary approach to inquiry. As student develop the genre-specific writing competencies, they also need to get a sense for the associated argumentative forums and dynamics of the disciplinary field in question. Thus, learning to write science entails moving beyond the formal practices of genre conventions to actively engaging with scientific evidence, knowledge, and concepts, and in the process, learning the social disciplinary standards and practices (Kelly et al. 2000).

Rhetorical analysis of professional scientific writing suggests that assessment of evidence is a cultural practice and can be investigated empirically. Textual analysis of students' writing and the assessment of this writing by populations with varied experience and knowledge in geological sciences therefore represents ways of understanding how evidence is used in science, and how it can be used by students in their own writing – issues we consider in our empirical study. Our two-part study is designed to identify the role of disciplinary-specific knowledge in assessing evidence, and subsequently, to make visible the practices of formulating evidence in geology writing. Both aspects of the study are oriented toward the goal of making explicit to students the disciplinary practices associated with formulating written evidence.

2. Educational Setting

This study was conducted in an introductory level oceanography course in a large public university and is part of a larger on-going educational ethnography of the course over the past six years (Kelly et al. 2000; Kelly & Takao, 2002). We chose to study this course for its emphasis on scientific writing in geological sciences. This undergraduate course satisfies a university writing requirement and emphasizes scientific writing. There are 2 one-hour and fifteen minute lectures per week and an additional weekly two hour laboratory meeting for this course. Students in the oceanography course are provided with the use of an interactive CD-ROM to access geological databases. This CD-ROM, 'Our Dynamic Planet' (Prothero, 1995), was created by the course professor. Data modules relating to plate tectonics are available through the still and moving graphics of this CD-ROM, such as earthquake locations and depths, volcanic locations, and the relative age of islands (Figure 1). More information about the CD-ROM may be found at http://oceanography.geol.ucsb.edu/.

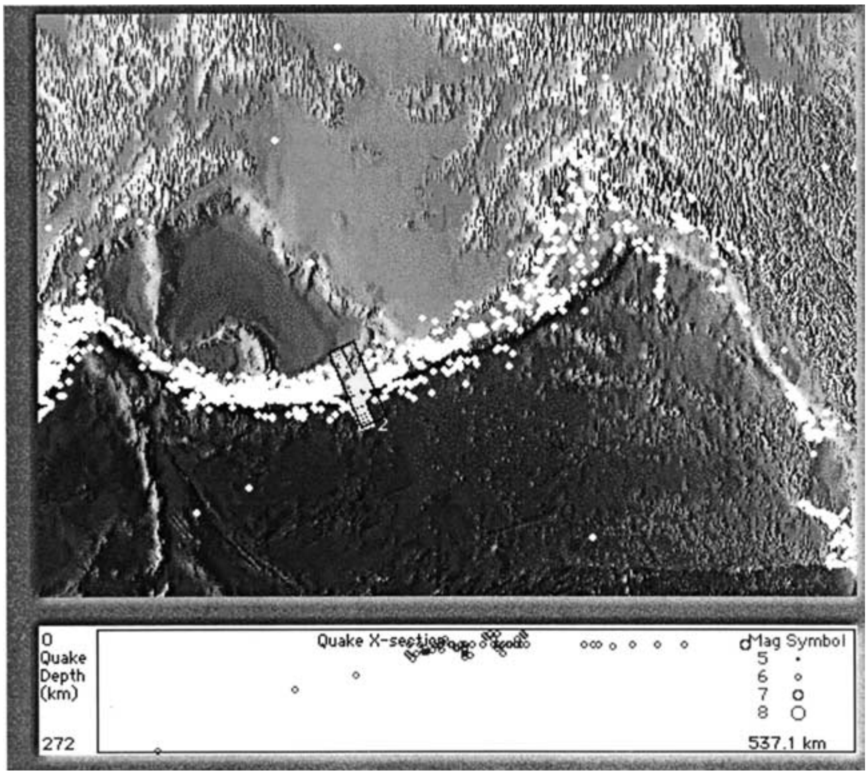

Figure 1. Map and earthquake depth profile from CD-ROM 'Our Dynamic Planet'.

Students use the information provided to them throughout the course via this CD-ROM, course lectures, laboratory sections, course textbook, and course reader to assist them in completing their mid-term scientific writing assignment. This assignment required students to write a technical paper characterizing several geographical areas using relevant geological data and to reconcile their findings with plate tectonic theory. The course reader provided an outline of the format for the technical paper and included descriptions and examples of each section of the paper. The technical papers followed a specific format defined by the instructor including the following sections: abstract, introduction, methods, observations, interpretations, conclusions, figures, references. The papers were approximately 6–10 pages of double-spaced text and including numerous data representations drawn from the multiple data sets provided by the CD-ROM.

3. Research Methods

Our study is organized into two parts. First, we considered the differences among and within three populations' assessments of students' scientific writing. This portion of the study sought to identify differences in interpretation of evidence by populations with differing subject matter knowledge. By comparing across the three populations (undergraduate students enrolled in oceanography, undergraduate students not enrolled in oceanography, course instructors), variations in the assessment of evidence could be made apparent. Second, we applied an argumentation analysis model to evaluate the argumentation structure of students' papers. We specifically focused on one paper rated high by the course instructors and on one paper rated low in order to make explicit the differences in uses of evidence. Through this two-part study we are interested in making the writing genre for geology accessible and available to students.

3.1. RESEARCH INTERVIEWING

The researchers (Takao and Kelly) and the course professor collaborated for the selection of the two papers used in this study. One paper was rated high in terms of the grade it received and by an independent review by the course professor. The other paper was rated low based on these same criteria. Both were chosen from a previous academic year with a similar writing assignment. These papers were chosen specifically to represent a strong contrast of argument formation in this technical genre. These papers were coded so the authors' identity and the papers' scores were unknown to the interview participants with the exception of the professor. The high scoring paper was coded with a red sticker and the low scoring paper was coded with a blue sticker. During the interviews the two papers were simply referred to as 'the blue paper' and 'the red paper'.

The three populations in our study had varying degrees of knowledge in geological sciences: subject matter experts (n = 4), the course professor and three teaching assistants; science students (n = 9), university students taking the introductory oceanography course; and non-science students (n = 8), university students who have not taken the introductory oceanography course and were in the undergraduate psychology pool of research subjects. These populations were selected because we were interested in learning how the level of discipline specific knowledge would influence the assessment of the quality of arguments in the students' papers. In other words, we wanted to gain an understanding of the knowledge drawn upon by subject matter experts to assess the quality of geological writing-knowledge that may not be available for the student population. In the first part of our study we attempted to elicit such knowledge.

The participants were individually interviewed and asked to compare the low scoring student paper with the high scoring student paper based on the following four criteria: overall opinion of the two papers, authors' use of evidence, authors' use of figures, and conclusions made by the authors. These four issues were identified as important factors to consider in scientific writing in a previous study (Kelly et al. 2000). The participants were asked to read both papers and then given a copy of our open-ended interview protocol (Patton 1990) to know what questions to expect during their interview. The interviews lasted between 30 minutes to 90 minutes and with the consent of the participants each interview was audio and video taped. Each interview was transcribed word for word. The transcripts of the interview audio and videotapes served as data for our analysis across and within the three populations' assessments of students' scientific writing.

In our initial analysis we open-coded (Emerson et al. 1995) the transcripts and multiple categories emerged from the data. Further in-depth analysis led to synthesis of the multiple categories into three categories of interviewees' positions regarding the four main issues: favorable, ambivalent, and unfavorable. We then reanalyzed the transcripts and focus-coded (Emerson et al. 1995) the data into these categories. We produced a chart demonstrating the participants' assessments of the low scoring and high scoring papers (Table I). As we describe in subsequent sections of this paper, analysis of the interviews revealed that the participants' reasoning patterns were rather ambiguous regarding their assessments of the differences between these two papers. This led us to make explicit the argumentation structures of the two papers, so that such differences could be made explicit to the student writers.

3.2. ARGUMENTATION ANALYSIS: EPISTEMIC LEVEL OF CLAIM

In the second part of our study, we applied an argumentation analysis model to examine the two papers used in the interviews. The argumentation analysis model used in this study was initially developed in a previous research study and substantially improved for this application (Kelly & Takao 2002). The argumentation model considers the formulation of evidence in two distinct but related ways. First, the model considers the epistemic level of the numerous claims comprising the overall argument. As described below, scientific argument generally is formed through claims of various degrees of generality. Second, the model takes into account the ways that evidence is built through an examination of the lexical cohesion (Halliday & Hasan 1976) of the various propositions comprising the argument. These two components work to characterize how evidential chains are established linking the data representations (e.g., earthquake profiles, volcano locations) to progressively more theoretical statements identifying geological features (e.g., trench, mountain range) and finally to the theoretical claims characterizing a geographical region in the overall plate tectonic model (e.g., convergent margin, subduction zone).

The epistemic level categories were based on textual analysis of scientific writing (Latour, 1987; Myers, 1997) and on Britton et al.'s (1975) transactional use of language in informative writing. An illustrative example is provided by Myers' (1997) rhetorical analysis of the evolution of two biology articles. In this analysis Myers identified how the authors needed to temper their knowledge claims through negotiations with their respective audiences. The scope of claims varied from relatively narrow foci (description of one species), to broader issues of interpretation (applicable to all species), to still broader issues of evolutionary processes, depend-

Table I. Summary of the participants' assessments of the low scoring and high scoring papers

Issue	Overall rating	Instructors $n = 4$		Oceanography students $n = 9$		Psychology students $n = 8$	
		Blue paper	Red paper	Blue paper	Red paper	Blue paper	Red paper
Overall opinion	Favorable	0	2	2	6	1	6
of quality	ambivalent	0	2	1	2	1	2
	unfavourable	4	0	6	1	6	1
Use of evidence	Favorable	0	2	2	6	1	6
	ambivalent	0	0	2	1	1	0
	unfavourable	4	0	3	0	4	1
Use of figures	Favorable	0	4	7	2	5	3
	ambivalent	0	0	2	2	2	3
	unfavourable	4	0	0	5	1	2
Conclusions	Favorable	0	2	4	6	2	4
	ambivalent	0	0	2	2	0	1
	unfavourable	3	0	3	0	4	1

ing on judgements referees and editors. Similarly, Latour's analysis of scientific writing suggested that scientists typically try to move rhetorically from low induction facts (i.e., very specific, grounded claims) to more generalized statements (i.e., theoretical claims) with respect to specific constructs of the relevant disciplinary-specific knowledge. In developing our initial argumentation analysis model (Kelly & Takao, 2002) we considered geological sciences-specific knowledge in relation to the students' argument structures. In this model there were six epistemic levels, from the most specific, grounded claims to progressively more general, theoretical claims. The model contains six levels from grounded data references to abstract theory. These levels are: representations of data, identification of topographical features, relational aspects of geological structures, data illustrations of geological theories or models, geological theory or model proposed by the author, description of geological processes and references to definitions, experts, and textbooks. An additional category 'PC' (Personal Comment) refers to statements in which the author offers meta-discursive comments to the reader. A summary of the epistemic level categories, definitions, and examples is provided in Table II.

The process of sorting each proposition into the relative epistemic levels was conducted as follows. The assignment for writing the technical paper included dividing the paper into preset sections: abstract, introduction, observations, interpretations, conclusions, and figures (Kelly et al. 2000). We analyzed the differences in epistemic levels of propositions comprising the sections labeled 'observations' and 'interpretations' of the two student papers since that was where much of the inferential work was done. First, we typed the text from the observations and the interpretations sections of the two student papers into computer files. Then we labeled each sentence with a proposition number for future cross-reference. Next, we sorted each proposition into an epistemic level based on the definitions of the epistemic categories. Then we placed the respective proposition number onto a semantic network (Figures 2 and 3). The semantic network indicates whether the sentence was from either the students' observations section (circles) or interpretations section (squares). The initial placement of each statement into the respective epistemic levels was completed by analyst 1 (Takao) and checked by analyst 2 (Kelly). We collaboratively reviewed all cases of disagreement until a consensus was reached. We placed the most specific claims on the bottom of the model and began our numerical system from this level. The numerical system we used was designed for referencing purposes in our rhetorical analysis. Although the numbers are in ordinal progression, they do not represent a quantitative measure of generality, nor should they be considered a measure of validity.

3.3. ARGUMENTATION ANALYSIS: LEXICAL COHESION

To further analyze the argumentation structure we evaluated links among statements across and within epistemic levels. These links reveal the connections (lexical cohesions, see Halliday & Hasan 1976) among students' propositions and

Table II. Definitions and examples of epistemic levels

Category	Definition	Examples
Epistemic Level VI:	General propositions describing geological processes and referencing definitions, subject-matter experts, and textbooks. The knowledge represented may not necessarily refer to data that is specific to the area of study.	‘It is known that at convergent plate boundaries there are deep earthquakes, volcanoes, and a deep trench a the boundary between the two plates (Ross, *Oceanograph*y, p. 52).’ Red (high-scoring) paper (Area 1, proposition #22)
Epistemic Level V:	Propositions in the form of geological theoretical claims or models specific to the area of study.	‘Many factors indicate that the Western Coast of South America is a subduction zone’. Blue (low-scoring) paper (Area 1, proposition #14)
Epistemic Level IV:	Propositions presenting geological theoretical claims or model illustrated with data specific to geographical area of study	‘We can see from figures 5,6, and 7 that the trench does in fact run the length of the island arc, telling us that the plate on the southern side is subducting under the plate on the northern side’. Red paper (Area 1, #23)
Epistemic Level III:	Propositions describing relative geographical relations amongst geological structures specific to the geographical area of study.	‘There also appears to be a large trench on the southern side of the island chain, extending the entire length of the island arc, about 3500 km in length’. Red paper (Area 1, #6)
Epistemic Level II:	Propositions identifying and describing topographical features of the geological structure specific to the geographical area of study.	‘The length of the trench is approximately 5900 kilometers and the width is 125 kilometers’. Blue paper (Area 1, #3)
Epistemic Level I:	Propositions making explicit reference to data charts, representations, locations, and age of islands or locating the geographical area of study.	‘The first area that is being examined is the Aleutian Island chain in the Northern Pacific Ocean. This area is just off the southwestern tip of Alaska (Fig. 1, area 1)’. Red paper (Area 1, #1–2)
Personal comment.	Statements by author to reader, often in the form of meta-discursive commentary. Not included on epistemic level scale.	‘There are many good features in this area that will lead to the conclusion on what type of boundary occurs here’. Red paper (Area 1, #3)

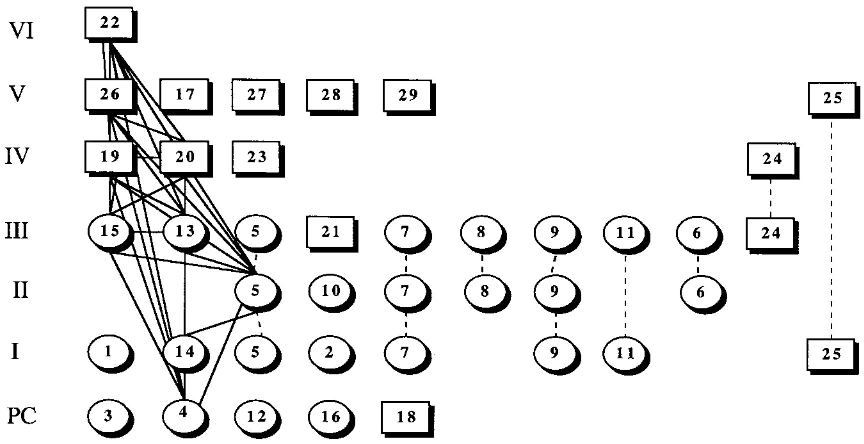

Figure 2. Distribution of students' propositions across epistemic levels of one geographical area for high paper. Numbers in square icons represent propositions by the student writer as 'interpretations'; numbers in round icons represent propositions defined by the student writer as 'observations'. Links across propositions are shown for only 'earthquake' evidence and are shown as solid lines connecting propositions. Dotted lines connect propositions of multiple epistemic levels.

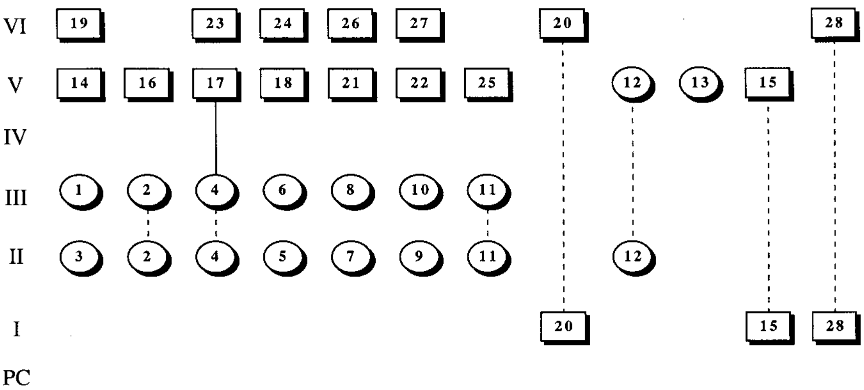

Figure 3. Distribution of students' propositions across epistemic levels of one geographical area for low paper. Numbers in square icons represent propositions by the student writer as 'interpretations'; numbers in round icons represent propositions defined by the student writer as 'observations'. Links across propositions are shown for only 'earthquake' evidence and are shown as solid lines connecting propositions. Dotted lines connect propositions of multiple epistemic levels.

are depicted on the semantic network as lines connecting propositions. We used the following five criteria to define the links: (a) Our first critierion was explicit links among statements across and within epistemic levels. We defined this as sentences using indexical phrases such as 'this' or 'its' in relation to a preceding or following statement. To illustrate this point consider the following propositions, 'The first area that is being examined is the Aleutian Island chain in the Northern Pacific

Ocean' and 'This area is just off the southwestern tip of Alaska (Figure 1 area 1 [world map indicating geographical area under study])' (high paper geographical area 1 propositions 1 and 2, respectively). Proposition 2 uses 'this' to explicitly reference the study area noted in proposition 1 and therefore they are linked to one another based on this first criterion. (b) Our second criterion was repetition of the same words or phrases within each area of the high and low scoring papers. We complied a list of words and phrases pertaining to geological content, such as 'trench', 'magma', 'island', 'arc', 'boundary', etc., based on the geological terms from the two papers. The students' written arguments were entered into a spreadsheet and searches were conducted compiling results for each repeated word. (c) Our third criterion was repetition of similar words or phrases regarding geological content, including synonyms, near-synonyms, variations of words (e.g., 'earthquake', 'earthquakes', 'quake', 'quakes', or 'ocean' and 'oceanic', etc.). (d) Our fourth criterion was superordinate propositions. We linked propositions that consisted of subordinate terms to propositions with superordinate terms. For example, the superordinate terms 'these three facts' from proposition 25 (high paper area 1) correlates with the subordinate three facts noted in propositions 19, 23, and 24. (e) Our fifth criterion was propositions that were sorted into more than one epistemic level. These were often in the form of compound sentences. On the semantic network a dotted-line link was used to connect the proposition numbers that are repeated across various epistemic levels. The repetition of proposition numbers on the semantic network may make proposition density appear seemingly higher than the actual true number of propositions being represented on the chart. Based on these five criteria, we calculated the amount of links across propositions.

4. Analyses and Findings

We conducted our analyses process of the two student papers in two phases. First, we analyzed the interview transcripts across and within the three populations' assessments of the high scoring and low scoring papers. Second, we applied our argumentation analysis model to further illustrate the differences between the high scoring and low scoring paper.

4.1. INTERVIEW RESULTS

In the first phase of our analysis, we considered the data from the interview transcripts. The interviews focused on four main issues: participants' overall opinion of the two papers, authors' use of evidence, authors' use of figures, and conclusions made by the authors. Several analyses lead to the synthesis of participants' positions regarding these issues into three categories: favorable, ambivalent, and unfavorable. We defined these categories as follows. "Favorable" referred to cases when the participant spoke positively of the issue under discussion. The following examples illustrate the participants favorable positions regarding their overall opinion of first the high scoring paper and second the low scoring paper, 'excellent paper its easy to read its concise it doesn't throw in a lot of irrelevant stuff its correct very excellent' (II;20;480),[2] 'it was really good I think that you'd get a good grade' (PSI 2;6;062). 'Ambivalent' referred to participant responses that were neither favorable or unfavorable towards the issue. For example, the following quotes are from the transcripts regarding the participants' overall opinion of the two papers. The first quote pertains to the high scoring paper and the second quote pertains to the low scoring paper, 'not a great paper but it's better written' (TAI 3;7;142), 'it [low scoring paper] was alright I don't think it got an A in my opinion' (OSI 9-1;1;012). 'Unfavorable' referred to participant responses that were negative regarding the issue at hand. For example, 'it stinks' (PSI 2;4;042) describes this participants' overall opinion of the high scoring paper, whereas 'there's a definite lack of understanding that's apparent' (II;13-14;312) refers to the overall opinion of this participant in regards to the low scoring paper. We point out that these categories do not necessarily imply that the participant favors one paper over the other with regards to each particular issue. The following two quotes serve to illustrate how a participant spoke favorably of the conclusions for both the high scoring paper and the low scoring paper, respectively: 'conclusion well thought out' (OSI 6-1;3;024), and 'conclusion really good' (OSI 6-1;5;056).

The interview results are presented in summarized form in Table I showing the participants' positions regarding four main issues for the low scoring and high scoring papers. Based on the interview transcripts and the summary chart, we found that more of the interview participants spoke favorably of the high scoring paper, than of the low scoring paper in terms of the following issues: participants' overall opinion of the paper (n = 14 high paper, n = 3 low paper), authors' use of evidence (n = 14 high paper, n = 3 low paper), and conclusions made by the authors (n = 12 high paper, n = 6 low paper). With regard to the authors' use of figures, the instructors (n = 4) all spoke favorably of this category for the high scoring paper and unfavorably for the low scoring paper. On the other hand, more of both the science students and the non-science students spoke favorably of the use of figures in the low scoring paper (7 science students, 5 non-science students) than that of the high scoring paper (2 science students, 3 non-science students). This analysis revealed that in terms of these four issues, the high scoring paper is regarded by the three populations as favorable in more instances than that of the low scoring paper. This was not perhaps a surprising result. However, a further analysis of the reasons offered for the relative differences in uses of evidence yielded different results.

Comparison of the participants' positions regarding the four categories indicated that in general the high scoring paper was favored relative to the low scoring paper by the three populations. Nevertheless, the reasons marshaled for this preference were ambiguous. Analyses of the transcripts revealed that the participants' were not particularly articulate in their reasoning for their relative positions regarding the four main issues. For example, a science student (OSI 3-1) states 'the second one [high scoring paper] is a lot stronger than the first one [low scoring paper].it's definitely well written'. The transcripts following this statement indicated the participant lacked clear reasons for the preference for the high scoring paper as 'stronger' and 'well written'. This was a typical response; participants made general overall assessments of the papers, yet they were not specific in pointing out examples from the papers to support their positions.[3] Similar results were found for the geology graduate student graders who were similarly ambiguous about the reasons for their perception of quality differences between the high scoring and low scoring papers. Therefore, to unpack the differences in uses of evidence across the two papers, we applied our argumentation model to the students' writing.

4.2. RESULTS OF ARGUMENTATION ANALYSIS

In the second phase of our analysis, we applied our argumentation analysis model to the high scoring and low scoring papers and noticed several differences between them. Our presentation of the results is in three parts. First, we present the argument formation as depicted by the argumentation analysis. The actual arguments made are presented and contrasts drawn. Second, we consider the overall argumentation structure in terms of the relative distribution of claims. Third, we present the results of analysis of the lexical cohesions tying claims together for each of the student arguments.

4.3. STUDENT SUBSTANTIVE ARGUMENTS

The argumentation analysis identified substantive differences in the students' uses of evidence. In the high scoring paper, theoretical claims (epistemic level V) were supported with data presented across multiple epistemic levels (I, II, II, IV). To illustrate this point we now present the argumentation structure for the high paper Area 1: Aleutian Islands. The author, 'Steve', used earthquake, volcanic, and topographical (i.e., trench) evidence to support his theoretical claim that a convergent subducting plate boundary exists at the Aleutian Island chain. These three sub-arguments were constructed with claims across various epistemic levels. For example, consider the earthquake evidence he used for that particular argument (mapped in detail in Figure 2). Steve first introduced the subject of earthquakes in proposition 4, a meta-discursive statement about the information he would present (sorted into epistemic level PC), 'The first thing that I looked at is the earthquake activity in the area'. He referenced several data representations (epistemic level I) to provide support for his theoretical claim, such as in proposition 14: 'Figures 8 and 9 [earthquake depth profiles] show the depth of the earthquakes along the profiles that were plotted'. He built his argument by identifying topographic features in propositions that were sorted into epistemic level II such as, 'There appears to be at least two hundred earthquakes directly along the chain of islands (Figure 2 [earthquake locations in Aleutian Island region])' (proposition 5). The next level of generality pertains to relational aspects of geological structures (epistemic level III) and the following proposition (#15) was sorted into that category, 'As observed in figure 8 [earthquake depth profile], the majority of the quakes are within 50 km of the surface, however there are about 20 recorded earthquakes that extend down the trench to around 252 km'. Steve's argument is further supported with proposition 19 which illustrated his theoretical claim with geological data from the study area (epistemic level IV), 'The abundance of earthquakes in the area and their locations is the first clue as to what type of tectonic process we are seeing'. For this particular argument, Steve also included a proposition which references information from the course textbook (epistemic level VI), 'It is known that at convergent plate boundaries there are deep earthquakes, volcanoes, and a deep trench at the boundary between the two plates (Ross, Oceanography, p. 52)'. These propositions (epistemic levels I, II, III, IV, and VI) served as evidence for his theoretical assertions in propositions 26 and 25 claiming the study area is a subduction zone (Figure 2).

The marshalling of evidence at multiple degrees of generality was a common pattern in Steve's writing. In a second geographical area, he continued to use evidence across epistemic levels to support his theoretical claims. Again he used earthquake evidence, this time to characterize the Mid-Atlantic ridge as a plate boundary. He introduced this line of reasoning as follows: 'The first step to determining any boundary type is by looking at the earthquake activity in the area' (epistemic level PC). He made explicit reference to several data representations (epistemic level I) and built his argument by identifying topographic features of the area (epistemic level II) such as in the following proposition: 'Along this stretch of the ridge, about 3500 km in length, there are about 100 seismic occurrences (Figure 3 [volcano locations in Aleutian Island region])'. Steve introduced more theoretically referenced data (epistemic level IV) to further support his theoretical claim that the Mid Atlantic ridge is a plate boundary (epistemic level V).

In comparison, the low scoring paper had multiple general geological theory statements which were left relatively unsupported by specific reference to and use of geological data. To illustrate this difference we turn to the argumentation structure of the low paper (mapped in detail in Figure 3), specifically considering the first geographical area identified by the student (South America). The author, 'Linda[4]' made the theoretical claim (epistemic level V) that 'Many factors indicate that the Western Coast of South America is a subduction zone' in proposition 14. Linda used earthquake evidence as one of these factors. Unlike the high paper, in which case we could trace Steve's use of earthquake evidence across epistemic levels, in this low paper the evidence exists in only one epistemic level from one proposition, which is proposition 17: 'The earthquakes along the trench also indicate a subduction zone'. When analyzing this paper we found it difficult to decipher the other 'factors' she intended as evidence. Therefore, at this particular geographical area and throughout her other arguments, Linda's theoretical claims were unsupported (Figure 3).

This general pattern continued for both the second and third geographical areas she considered. In the case of the second geographical area, Linda presented earthquake evidence to support her theoretical claim (epistemic level V) that: 'The Himalayas were formed by a continental/continental plate collision (see Figure 4 [map and elevation profile in Himalayan mountain range region])'. Earthquake evidence was referenced in two propositions classified at epistemic levels II and III. She did not make use of specific data representations, nor offer other substantial basis for the claims regarding earthquakes. Thus, while she did reference earthquakes in a vague manner, the relationship to the theoretical assertion regarding the Himalayas to the relevant earthquake data was left untethered. For the third geographical area investigated by Linda, she discussed volcanic activity as evidence for her theoretical assertion (epistemic level V) that 'The Hawaiian Island Chain was formed by a hot spot'. However, this reference to volcanic activity was in only one proposition at epistemic level II. Other references to volcanic information were found in multiple theoretical propositions that were sorted into epistemic level V (5 propositions) and VI (3 propositions) which essentially do not use data from the study area as evidence to support her theoretical claims. In sum, Linda made a large number of unsupported theoretical assertions. This general characteristic becomes more obvious through a comparison of the distribution of claims across epistemic levels for each of the writing samples.

4.4. DISTRIBUTION OF CLAIMS ACROSS EPISTEMIC LEVELS

The overall distribution of claims across epistemic levels similarly showed significant differences between the two papers. The propositions from the high scoring paper were distributed across the various epistemic levels, with many mid-level claims (epistemic levels II, III, and IV; as summarized in the third column of Table 3). A majority of the propositions from the low scoring paper were classified into epistemic levels V or VI (Area 1: 17 of 28; Area 2: 11 of 17; Area 3: 18 of 24). The distribution of propositions is related to our first finding discussed above. As previously noted, the high paper used evidence across multiple epistemic levels so that we would see these propositions distributed across our semantic network. On the other hand, the low scoring paper has a more skewed distribution of propositions which relates to our previous finding that supporting evidence was generally absent as there were few lower-inference propositions aligned with particular theoretical assertions. Interestingly, the low scoring paper had no propositions sorted into epistemic level IV (see Table III) – propositions that generally tie theory to the specific geological features of the region in question. In addition, theoretical assertions of epistemic level V in the high scoring paper were proposed strictly in their 'interpretations' section, whereas the low scoring paper had theoretical statements from both 'observations' and 'interpretations' sections. This may indicate that Linda was unclear about the level of inference permitted for 'observations'.

Table III. Number of propositions, links, link density, and distribution of propositions across epistemic levels for sampled student writing

Sample of student Writing	Number of propositions	Distribution of propositions* across epistemic levels		Number of links across propositions	Number of links per proposition
		EL	#		
High paper, geographical area 1	29	VI	1	245	8.45
		V	6		
		IV	4		
		III	10		
		II	6		
		I	8		
High paper, geographical area 2	22	VI	3	222	10.1
		V	4		
		IV	1		
		III	10		
		II	3		
		I	8		
Low paper, geographical area 1	28	VI	7	129	4.61
		V	10		
		IV	0		
		III	7		
		II	8		
		I	3		
Low paper, geographical area 2	17	VI	2	48	2.82
		V	9		
		IV	0		
		III	2		
		II	3		
		I	1		
Low paper, geographical area 3	24	VI	5	154	6.42
		V	13		
		IV	0		
		III	3		
		II	4		
		I	2		

*Compound sentences may contain multiple propositions sorted into more than one epistemic level.

4.5. ARGUMENT COHERENCE THROUGH LEXICAL COHESIONS

Scientific arguments not only need to marshal evidence through the use of claims at various levels of generality, but these claims need to be tied together into a coherent whole. Our next analysis was a linguistic examination of lexical cohesions, applying the five criteria described in the previous 'research methods' section. We found that the highly rated paper had a greater number of links for each geographical area presented, respectively, in comparison to the low paper (see Table III). For the high paper geographical area 1 there are 245 links and for area 2 there are 222 links. However, for the low paper, in geographical area 1 there are 129 links, area 2 has 48 links, and area 3 has 154 links. Taking into account the total number of propositions, the general pattern is maintained: that is, the number of links per proposition was greater for the higher scoring paper for each geographical area studied (8.45 and 10.1 versus 4.61, 2.82, 6.42). Due to the large number of lexical cohesions in each paper, the links across epistemic levels for only one line of evidence are presented in Figures 2 and 3: The 'earthquake' line of evidence with the cohesive links is shown for each paper making visible the differences between the theory-data ties in each.

Although it may seem obvious that links between statements are key components for building a strong argument, this point was not articulated during the interviews. For instance, the research interviews did not indicate that the high scoring paper tied theoretical claims to supporting evidence, a characteristic made evident through argumentation analysis. In other words, our model did allow us to draw out some differences between the papers which were not articulated in the interviews, even for geology graduate student teaching assistants who served as graders of the undergraduate papers. These findings allowed us to identify components of the students' arguments from a normative point of view and to pose pedagogical implications which we discuss in the subsequent section.

5. Discussion and Educational Implications

Our results indicated that while evidential quality was generally recognized by the oceanography course instructors, the oceanography students, and even students not enrolled in this course, explicit reasons for preferences regarding the uses of evidence were minimal. This suggested a need for identifying and documenting how evidence is marshaled in scientific writing. Drawing from theories of rhetoric of science and argumentation more generally, we designed and applied an argumentation analysis model that considered evidence formation in two substantive ways. First, the model considered the epistemic level of claim of each of the propositions forming the overall argument. This allowed us to make explicit the ways that theoretical claims (e.g., characterizing a subduction zone) were supported using geological data (e.g., earthquakes depth profiles) for a given geographical region. Second, the model considered the ways in which claims across epistemic levels were linked together through lexical cohesion (e.g., collocation, repetition). This process identified how student writers organized claims together to form an overall argument. The study has implications for the writing and assessing of evidence and for discussion of uses of evidence in educational settings.

Specific implications can be drawn for instruction in writing in this genre, for both students and teachers. The task of formulating scientific arguments requires abstraction from specific data to make theoretical claims. However, this abstraction process requires intermediate steps, where the level of claim is more than description of data representations, but not as general as the overall theoretical assertion of the main argument. This requires student understanding certain features of scientific writing. First, students may need experience assessing the level of claim in their own and others' writing. Experience assessing the extent of the generality of claims may give students insight into the types and extent of evidence required for a given claim. Second, the facts need to be 'stacked' from most descriptive to most general through progressive abstraction (Latour 1987). Building a cohesive argument requires making connections across the levels of abstraction. Assessing the relationship of different parts of an argument, and how evidence is formulated for each part, may give students better insight into ways of unpacking the presented evidence. This may have significant implications for science-technology-society issues, where there are often ethical implications mixed with scientific information. Finally, for teachers of scientific writing the use of this argumentation model identified previously invisible features of the writing task. Considering the ways that claims are made in an explicit manner may give instructors insight into their own tacit knowledge embedded in the writing process.

Science educators have been deeply concerned about the importance of rationality (Siegel 1991; Strike & Posner 1992) and reason in science (Loving 1997; Matthews 1997). Furthermore, the importance of theory for organizing learning and understanding the value of empirical evidence has been identified as central to scientific activity and as a normative goal for science education (Duschl 1990). In this study we sought to bring a naturalistic point of view to understand evidence formation (Giere 1999) for a particular genre of scientific writing, and thus begin to demystify the rhetorical and linguistic features that make scientific arguments obscure for student readers and writers (Halliday & Martin 1993). The move to the empirical investigation of epistemological issues in science education follows the suggestion of a recent review of studies in the nature of science which documented the preponderance of survey questionnaires (Kelly et al. 1998; Lederman et al. 1998). From our point of view, understanding the nature of science includes not only understanding about scientific practices, but the ability to engage in these practices in an effective manner. Therefore, understanding the uses and limitation of large scale data sets through writing represents one way for students to engage in the epistemic practices of scientific communities. The analysis of the uses of evidence in writing depicted in our argumentation analysis model represents one way to develop the more open-ended, qualitative approaches to the assessment of students' understanding suggested by Lederman et al. (1998). This mode of assessment of written arguments, however, is relatively time consuming in comparison to that of multiple-choice or short answer assessments. In addition, the amount of time it takes to complete such an assessment increases when, in addition to evaluating the structure of the argument and the quality of the claims, the evaluator attempts to assess the chains of reasoning across claims. Nevertheless, the theoretical issues raised here can be translated into pedagogy. For example, our argumentation analysis model of epistemic levels has been translated by the course instructor as heuristic for geological-specific scientific writing. This heuristic requests that students place propositions from the two papers analyzed in this study into different epistemic levels. The process of sorting propositions is designed to provide opportunities for the student writers to get a sense for the ways empirical and theoretical claims layering into an evidence-based argument. Further details are explained in detail in Takao et al. (2002). Writing has an unique place in shaping scientific knowledge (Bazerman 1988; Myers 1997) and offers students epistemological tools to develop understandings of evidence formation (Prain & Hand 1999). Through detailed analysis of argument science educators can develop more thorough understandings of scientific reasoning as relevant to student understanding of epistemic practices.

Acknowledgements

The research reported in this article was supported by grants from the National Science Foundation (NSF OCE 9809579, NSF DUE 9952491). The data presented, the statements made, and the views expressed are solely the responsibility of the authors. We would like to thank Julie Bianchini for her helpful comments on an earlier draft of this paper.

Notes

[1] To avoid confusion, we would like to note that in the field of the rhetoric of science, rhetoric is not, as in common parlance, opposed to argument. Rather, this field of study (Perelli 1989; Gross 1990; Bazerman 1988) investigates empirically how scientific discourse is used to persuade a relevant audience of the merits of an argument. Indeed, it is our view that the most persuasive arguments in science and science education are those that marshal evidence for the audience in question taking into consideration the norms, expectations, knowledge, social practices, etc. of the relevant epistemic community.

[2] Each quote from the interviews is indexed by interviewee role (II = instructor interview; TAI teaching assistant interview; OSI# oceanography student interview and student number; PSI# psychology student interview and student number); videotape number; and transcript line number.

[3] We readily acknowledge that research interviewing is an imperfect method for gleaning information and that the lack of thorough discussion of specific reasons for preferring one writing sample over another may have been due in part to the interviews themselves (Kvale 1996; Mishler 1986; Spradley 1979). Specifically, interviewees may have had and used knowledge relevant to their differentiation of the merits of the papers that was not distinguished through the interview process. Furthermore, we acknowledge that the selection of category topics and distribution of such information into these categories was determined by the authors' analytic decisions. Despite these

methodological limitations the information gleaned from the interviews was useful in two respects. First, it showed, minimally, that if there was disciplinary-specific knowledge brought to bear on the analysis of the student papers, this knowledge is not easily identifiable. Second, the ambiguities of the interview responses provided a rationale for further development of the argumentation analysis methods applied to the student writing.

[4] The two papers were chosen without the reviewers knowing the students' gender. As it turned out, the strong paper happened to be written by a male student, and the weak paper by a female. However, in a larger sample there were no gender differences as measured argumentation strength or in student grades (see Kelly & Takao, 2002).

References

Atkinson, D.: 1999, *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London 1675–1975*, Lawrence Erlbaum, Mahwah, NJ.

Bazerman, C.: 1988, *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*, University of Wisconsin Press, Madison, WI.

Bezzi, A.: 1999, 'What is This Thing Called Geoscience? Epistemological Dimensions Elicited with the Repertory Grid and Their Implications for Scientific Literacy', *Science Education* **83**(6), 675–700.

Britton, J., Burgess, T., Martin, N., McLeod & Rosen, H.: 1975, *The Development of Writing Abilities (11–18)*, Macmillan, London.

Carlsen, W.S.: 1997, 'Never Ask a Question if You Don't Know the Answer: The Tension in Teaching Between Modeling Scientific Argument and Maintaining Law and Order', *Journal of Classroom Interaction* **32**(2), 14–23.

Driver, R., Newton, P. & Osborne, J.: 2000, 'Establishing the Norms of Scientific Argumentation in Classrooms', *Science Education* **84**(3), 287–312.

Duschl, R.A.: 1990, *Restructuring Science Education: The Importance of Theories and Their Development*, Teacher's College Press, New York.

Emerson, R.M., Fretz, R.I. & Shaw, L.L.: 1995, *Writing Ethnographic Fieldnotes*, The University of Chicago, Chicago, IL.

Jimenez-Aleixandre, M.P., Rodriguez, A.B. & Duschl, R.A.: 2000, ' "Doing the Lesson" or "Doing Science": Argument in High School Genetics', *Science Education* **84**(6), 757–792.

Giere, R.: 1999, *Science Without Laws*, University of Chicago Press, Chicago.

Gieryn, T.F.: 1999, *Cultural Boundaries of Science: Credibility on the Line*, University of Chicago Press, Chicago.

Goodwin, C.: 1995, 'Seeing in Depth', *Social Studies of Science* **25**(2), 237–274.

Grandy, R.E.: 1997, 'Constructivisms and Objectivity: Disentangling Metaphysics from Pedagogy', *Science & Education* **6**(1–2), 43–53.

Gross, A.: 1989, *The Rhetoric of Science*, Harvard University Press, Cambridge, MA.

Halliday, M.A.K. & Hasan, R.: 1976, *Cohesion in English*, Longman, London.

Halliday, M.A.K. & Martin, J.R.: 1993, *Writing Science: Literacy and Discursive Power*, University of Pittsburgh Press, Pittsburgh.

Harris, R.A.: 1997, 'Introduction', in R.A. Harris (ed.), *Landmark Essay on the Rhetoric of Science: Case Studies*, Erlbaum, Mahwah, NJ, pp. xi–xlv.

Kelly, G.J., Chen, C. & Crawford, T.: 1998, 'Methodological Considerations for Studying Science-in-the-Making in Educational Settings', *Research in Science Education* **28**(1), 23–49.

Kelly, G.J., Chen, C. & Prothero, W.: 2000, 'The Epistemological Framing of a Discipline: Writing Science in University Oceanography', *Journal of Research in Science Teaching* **37**(7), 691–718.

Kelly, G.J., Druker, S. & Chen, C.: 1998, 'Students' Reasoning About Electricity: Combining Performance Assessments with Argumentation Analysis', *International Journal of Science Education* **20**(7), 849–871.

Kelly, G.J. & Takao, A.: 2002, 'Epistemic Levels in Argument: An Analysis of University Oceanography Students' Use of evidence in Writing', *Science Education* **86**(3), 314–342.

Keys, C.W.: 1999, 'Revitalizing Instruction in Scientific Genres: Connecting Knowledge Production with Writing to Learn in Science', *Science Education* **83**(2), 115–130.

Keys, C.W.: 2000, 'Investigating the Thinking Processes of Eighth Grade Writers During the Composition of a Scientific Laboratory Report', *Journal of Research in Science Teaching* **37**(7), 676–690.

Kuhn, D.: 1992, 'Thinking as Argument', *Harvard Educational Review* **62**(2), 155–178.

Kuhn, D.: 1993, 'Science as Argument: Implications for Teaching and Learning Scientific Thinking', *Science Education* **77**(3), 319–337.

Kvale, A.: 1996, *InterViews*, Sage, Thousand Oaks, CA.

Latour, B.: 1987, *Science in Action: How to Follow Scientists and Engineers Through Society*, Harvard University Press, Cambridge, MA.

Lederman, N.G., Wade, P.D. & Bell, R.L.: 1998, 'Assessing the Nature of Science: What is the Nature of Our Assessments?', *Science & Education* **7**(6), 595–615.

Lemke, J.L.: 1990, *Talking Science: Language, Learning, and Values*, Ablex, Norwood, NJ.

Loving, C.: 1997, 'From the Summit of Truth to its Slippery Slopes: Science Education's Journey Through Positivist-Postmodern Territory', *American Educational Research Journal* **34**(3), 421–452.

Matthews, M.R.: 1997, 'James T. Robinson's Account of Philosophy of Science and Science Teaching: Some Lessons for Today From the 1960's', *Science Education* **81**(3), 295–315.

Mishler, E.G.: 1986, *Research Interviewing: Context and Narrative*, Harvard University Press, Cambridge.

Myers, G.: 1997, 'Texts as Knowledge Claims: The Social Construction of Two Biology Articles', in R.A. Harris (ed.), *Landmark Essay on the Rhetoric of Science: Case Studies*, Erlbaum, Mahwah, NJ, pp. 187–121.

Newton, P., Driver, R. & Osborne, J.: 1999, 'The Place of Argumentation in the Pedagogy of School Science', *International Journal of Science Education* **21**(5), 553–576.

Patronis, T., Potari, D. & Spiliotopoulou, V.: 1999, 'Students' Argumentation in Decision Making on a Socio-Scientific Issue: Implications for Teaching', *International Journal of Science Education* **21**(7), 745–754.

Patton, M.G.: 1990, *Qualitative Evaluation and Research Methods*, Sage, Newbury Park, CA.

Prain, V. & Hand, B.: 1999, 'Students Perceptions of Writing for Learning in Secondary School Science', *Science Education* **83**(2), 151–162.

Prelli, L.J.: 1989, *A Rhetoric of Science: Inventing Scientific Discourse*, University of South Carolina, Columbia, SC.

Prothero, W.A.: 1995, 'Taming the Large Oceanography Class', *Journal of Geological Education* **43**, 497–506.

Richmond, G. & Striley, J.: 1996, 'Making Meaning in Classrooms: Social Processes in Small-Group Discourse and Scientific Knowledge Building', *Journal of Research in Science Teaching* **33**(8), 839–858.

Russell, T.: 1983, 'Analyzing Arguments in Science Classroom Discourse: Can Teachers' Questions Distort Scientific Authority?', *Journal of Research in Science Teaching* **20**(1), 27–45.

Schwegler, R.A. & Shamoon, L.K.: 1991, 'Meaning Attribution in Ambiguous Texts on Sociology', in C. Bazerman & J. Paradis (eds.), *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*, University of Wisconsin Press, Madison, pp. 216–233.

Siegal, H.: 1991, 'The Rationality of Science, Critical Thinking, and Science Education', in M. Matthews (ed.), *History, Philosophy, and Science Teaching: Selected Readings*, OISE Press, Toronto, pp. 45–62.

Spradley, J.P.: 1979, *The Ethnographic Interview*, Harcourt Brace Jovanovich, Fort Worth.

Strike, K.A. & Posner, G.J.: 1992, 'A Revisionist Theory of Conceptual Change', in R. Duschl & R. Hamilton (eds.), *Philosophy of Science*, *Cognitive Psychology*, *and Educational Theory and Practice*, SUNY Press, Albany, NY, pp. 147–176.

Swales, J.M.: 1990, *Genre Analysis*, University Press, Cambridge.

Takao, A.Y., Prothero, W. & Kelly, G.J.: 2002, 'Applying Argumentation Analysis to Assess the Quality of University Oceanography Students' Scientific Writing', *Journal of Geoscience Education* **50**(1), 40–48.