

## Analyzing the Quality of Argumentation Supported by Personally-Seeded Discussions

Douglas B. Clark

Victor D. Sampson

College of Education  
Division of Curriculum and Instruction  
Arizona State University  
Payne Hall, PO Box 870911  
Tempe, Arizona, 85287-0911  
Tel: (480) 491-7921, Email: dbc@asu.edu

**Abstract.** Several researchers have shown that student participation in discourse paralleling that of scientific communities is critical to successful science education. This study focuses on supporting scientific argumentation in the classroom through a personally-seeded online discussion system. Students use an online interface to build principles to describe data they have collected. These principles become the seed comments for the online discussions. The software sorts students into discussion groups with students who have built different principles so that each discussion group can consider and critique multiple perspectives. We outline a methodology for (a) coding the individual comments in terms of epistemic operation, grounds, and content normativity and (b) parsing and assessing overall argumentation structure of the oppositional episodes. This study therefore contributes to the research literature both in terms of scaffolding and assessing student argumentation in online asynchronous forums.

**Keywords:** Education, Conversation Analysis, Argumentation, Discussions, Secondary School

### INTRODUCTION

Research shows that students' participation in discourse paralleling that of scientific communities is critical to successful science education (e.g., Lemke, 1990; Rosebery, Warren, & Conant, 1992; Schauble, Glaser, Duschl, Schulze, & John, 1995). Argumentation is a genre of discourse crucial to the practice of science (Driver, Newton, & Osborne, 2000; Kuhn, 1993; Lemke 1990; Siegel, 1995; Toulmin, 1958), and much of science involves dialectical and rhetorical argumentation (Latour & Woolgar, 1986; Longino, 1994). This study focuses both on scaffolding and assessing scientific argumentation in the classroom through a customized online discussion system.

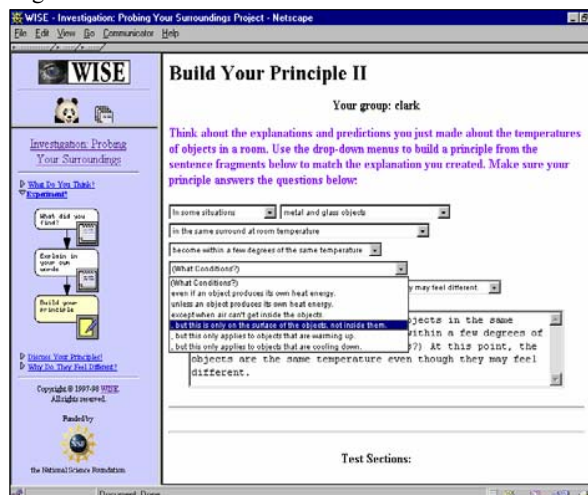
### Computer-based supports for argumentation: Personally-seeded discussions

Structured environments have been built to support scientific argumentation, discourse, and knowledge refinement. Some of these environments like Collaboratory Notebook (Edelson, Pea, & Gomez, 1996), CaMILE (Guzdial, Turns, Rappin, & Carlson, 1995), and Knowledge Forum/CSILE (Scardamalia, Bereiter, & Lamon, 1994) are learning environments unto themselves that focus heavily on knowledge collection and building. Others, like SpeakEasy (Hoadley, Hsi, & Berman, 1995), Sensemaker (Bell, 1997), and the BGUILE data reporting section (Tabak, Smith, Sandoval, & Reiser, 1996), are part of larger inquiry environments. In addition to these specialized environments, basic online threaded asynchronous forums in which discussions are held have also been shown to be effective in supporting classroom-based discourse (Collison, Elbaum, Haavind, & Tinker, 2000; Salmon, 2000).

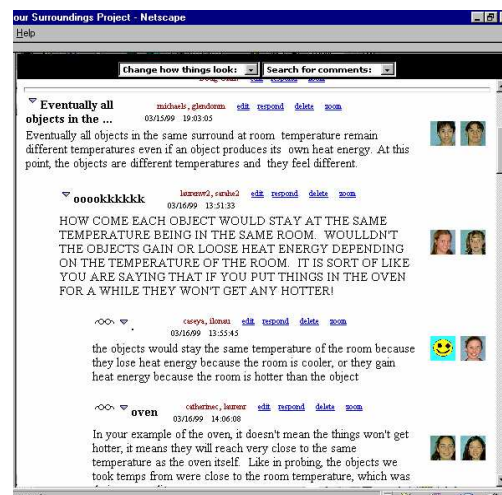
Whereas the online environments detailed above focus either on sharing information or on preparing arguments for presentation, our personally-seeded discussion system focuses specifically on engineering and supporting scientific argumentation within classroom discourse. Personally-seeded discussions (a) help students synthesize a principle to describe data that they have collected or found in light of other evidence from their classroom and homes, (b) create groups of students who have created different principles to describe the data, (c) facilitate online discourse among the students where they critique each other's principles in light of the evidence and work toward consensus through scientific argumentation based on the evidence, and (d) provide students with models of productive scientific argumentation.

The personally-seeded discussion system analyzed in this study is embedded in an online inquiry project. After collecting data, students create principles to describe patterns in the data. Research on students' initial conceptions about heat and temperature (Clark, 2000, 2001; Lewis, 1996; Linn & Hsi, 2000) and earlier thermodynamics curriculum development (Lewis, Stern, & Linn, 1993) form the foundation of a new Web-based

principle-builder interface that allows students to construct scientific principles from a set of predefined phrases and elements (Figure 1). After students create their principles, the project software places the students in electronic discussion groups with students who have constructed different explanatory principles. A screenshot of a portion of an asynchronous discussion within the thermodynamics project from this study is included in Figure 2.



**Figure 1.** Students use the principle-builder to construct scientific principles that become initial discussion comments



**Figure 2.** An example of a personally-seeded discussion

The student-constructed principles appear as the seed comments in the discussions. The discussions develop around the different perspectives represented in the seed comments, ideally through a process of comparison, clarification, and justification. As part of this process, the students are required to support their assertions and claims with evidence from their labs and other experiences. This process attempts to elicit self-explanation by helping students focus other students' attention on possible inconsistencies in their explanations and on reasoning, plausibility, completeness, and other attributes of "good explanations." In these discussions, all students and their ideas become critical resources with the common goal of refining individual student ideas.

## Analysis of student argumentation

To make judgments about argumentation quality, researchers over the last decade have developed several different methods to identify the essential features of an argument. These methods have been used to examine the structure of student arguments in small group conversations (Forman, Larreamendy-Joerns, Stein, & Brown, 1998; Kelly, Druker, & Chen, 1998; Resnick, Salmon, Zeitz, Wathen, & Holowchak, 1993) and in writing (Bell & Linn, 2000; Kelly & Takao, 2002). To date, most of these investigations of student discourse have relied heavily on Toulmin's (1958) model for argument structure in one way or another. In these studies, emphasis is placed on the identification of the structural features of arguments (e. g., claims, data, warrants, backings, and qualifiers) and the process of argumentation, especially in terms of how students provide warrants for claims. Such approaches seek to identify the absence or presence of the components of argument and use this information to assess argumentation quality. Structural analyses of student arguments contribute to our understanding of how students assimilate the desired practices of argumentation (Driver, Newton, & Osborne, 2000) and provide a great deal of information about the form and type of reasoning that students use when they construct arguments based on their everyday experiences (Simon, Osborne, & Erduran, 2003).

Although structural analyses of student argumentation is important in understanding students' reasoning when they construct arguments, these analyses should also include judgments of the quality of arguments in terms conceptual adequacy (Sandoval, 2003; Schwarz, Neuman, Gil, & Iiya, 2003; Zohar & Nemet, 2002). For instance, Zohar and Nemet (2002) found that although a majority of students are able to construct simple scientific arguments in the context of a unit on genetics, only a small minority included correct, specific scientific knowledge in their arguments. In our own research, we have found that students often include grounds which incorporate real world examples or data from prior labs when making claims or rebuttals. For example:

Why do you disagree? Remember when we did the potato lab? Even the things that were well insulated would eventually reach room temperature. Also remember how circuits use copper wire because we know that copper is a good conductor! Obviously the good conductor would reach room temperature first as opposed to bad conductors.

From the perspective of argumentation structure, a rebuttal like this is excellent in its use of data and warrants. However, even though the students raise important data from a prior lab activity, they confuse electrical and thermal conductivity in their real-world example. Normative conceptual content unfortunately does not always accompany desirable argumentation structure. As science educators strive to help students develop the cognitive skills required to construct, evaluate and defend scientific arguments, analytic methods must be developed to assess the quality of argumentation structure as well as the quality of the scientific content.

## RESEARCH QUESTIONS

In addition to developing our personally-seeded discussion environment, we are concurrently developing a method to assess student argumentation based on the epistemic operation of the comments that students make, the grounds students use, and the conceptual quality of their arguments. This current study asks three questions: (1) What is the nature of the arguments constructed by students in our personally-seeded environment in terms of epistemic operation, grounds, and conceptual quality? (2) Is there a relationship between the structural quality of the episode in which a comment occurs and the quality of the comment's grounds or content? (3) How robust and reliable is this new coding method?

## METHODS

This study incorporates and augments the coding schemes developed by several researchers to analyze the structure of student argumentation (e. g., Jimenez-Aleixandre, Rodrigues, & Duschl, 2000; Simon, Erduran, & Osborne, 2002). We first outline the methods we used to code each individual comment in terms of epistemic operation, quality of grounds, and quality of subject matter. Following the coding discussion for the individual comments, we outline our method of parsing and scoring the oppositional episodes within which the comments take place.

### Participants

Eight randomly chosen online discussions involving a total of 84 students are analyzed from four classes of eighth grade students who completed the project during one semester under the supervision of an experienced teacher who has worked extensively with the researchers. The public school is located in a diverse city and has an even distribution of boys and girls. The classes are typical 8th grade physical science classes, labeled neither "honors" nor "remedial." Each online discussion involves approximately five pairs of students. Students work on the project in pairs over the course of six class periods (five hours in total). The discussions begin at the start of the fourth class period and extend through the end of the fifth class period. To represent multiple perspectives, the software assigns student pairs to discussions with students who have created different principles, as discussed above.

### Coding Individual Comments

Students make a total of 334 comments in the 8 discussions. All comments are coded in light of the parent comment to which they reply, which means that the comments are coded in context rather than as individual statements.

#### *Coding the epistemic operation of a comment*

We code the epistemic operation of a comment in terms of the comment's role (or intended role) in a co-constructed dialogic argument. We code each comment in relation to the actual parent comment to which it responds to avoid ambiguity in terms of references within a comment. The coding scheme is outlined in Table 1 below. These codes take into account comments that are part of the actual argumentation, meta-organizational comments within the discussion that are not truly part of the argumentation but that help organize the interaction, and the occasional off-task interactions.

Table 1. Coding scheme for Epistemic operation of individual comments.

<i>Claim (CM)</i> : A hypothetical statement or a seed-comment principle made by a group of students.
<i>Counter-Claim (CC)</i> : A hypothetical statement or a principle made by a group of students that is different from and (does not attack in any way) the seed claim or parent comment made by another group. This code is only assigned when a comment does not focus on any aspect of the thesis of the comment it replies to; instead it offers an entirely new interpretation of the phenomena.
<i>Rebuttal Against Grounds (RAG)</i> : An attack on, or disagreement with, the grounds (evidence, explanations, qualifiers, or backing) used by another group to support or justify their comment. Comments in this category include: (1) using a rhetorical question as a way to question the validity of the grounds used by the other group, (2) statements that attempts to limit the conditions that evidence can be used, (3) a reinterpretation of the grounds used by the other group, (4) disagreement with the way empirical data was gathered, (5) disagreement with the way data is used, or (6) disagreement with the accuracy of the empirical data.
<i>Rebuttal Against Thesis (RAT)</i> : An attack on or disagreement with the thesis (or a specific part of the thesis) of another group's comment (claim or rebuttal) that does not attack the support (grounds) used by the other group. This category includes: (1) using a rhetorical question as a way to question the validity of the claim used by the other group, (2) asking a specific question about some portion of a comment where the intent is not to clarify the meaning but rather to question validity or accuracy, (3) correcting a specific aspect of another groups claim or rebuttal but not the grounds, (3) comments that express disagreement with the thesis of another group's comment and then offer a new claim.
<i>Support of a Comment (SC)</i> : A statement used to support the truth or accuracy of the previous claim or rebuttal. This category includes statements that (1) voice agreement with a comment (2) rewords the previous comment (3) adds additional grounds in support or (4) expands the comment.
<i>Query about Meaning (QUM)</i> : A comment that asks for clarification of an earlier comments (e.g., "What do you mean when you say...?" or "I don't understand what you are saying?"). These comments question the meaning of a statement rather than the accuracy of the statement. Therefore, rhetorical questions used as a way to question the validity of a claim or grounds are not included in this category.
<i>Clarification of Meaning (CLM)</i> : A comment made by a group of students to clarify (restate in a new way) a previous comment. The purpose of these comments is to clarify the meaning of a statement in response to a query (about meaning) rather the supporting the accuracy of a statement
<i>Clarification in response to a Rebuttal (CLR)</i> : This code is assigned to comments that are used to strengthen a position (in terms of accuracy or validity) in response to a rebuttal without attacking the rebuttal or grounds made by another group.
<i>Change of Claim (CH)</i> : A comment made by a group of students that indicates that (1) they have changed their original claim or (2) changed their viewpoint, or (3) have made a concession in response to comments (claims or rebuttals) made by another.
<i>Organization of Participants (OP)</i> : A comment that (1) reminds other participants to participate, (2) asks others for feedback, (3) has a metacognitive aspect (e. g. "Do we all agree?"), (4) attempts to change the way someone else in the discussion is participating.
<i>Off Task (OT)</i> : Comments that are not about the topic (e.g., "Nice haircut, John!").

### Coding the grounds of a comment

Once these base codes are assigned to characterize epistemic operation, the grounds are coded using the flow chart in Figure 3. Grounds include data, warrants, and backings (e.g., "The metal chair felt different but it was room temperature in our experiment"). Erduran et al. (Erduran, Osborne, & Simon, 2004; Simon et al., 2003) collapsed this category because of pragmatic challenges in reliably differentiating data, warrants, and backings in student transcripts. Rather than attempting to differentiate between data warrants and backing we classified the grounds of a comment as either: no grounds (level 0), using an explanation as grounds (level 1), using evidence as grounds (level 2), and coordinating multiple pieces of evidence or multiple connections between ideas in the evidence (level 3).

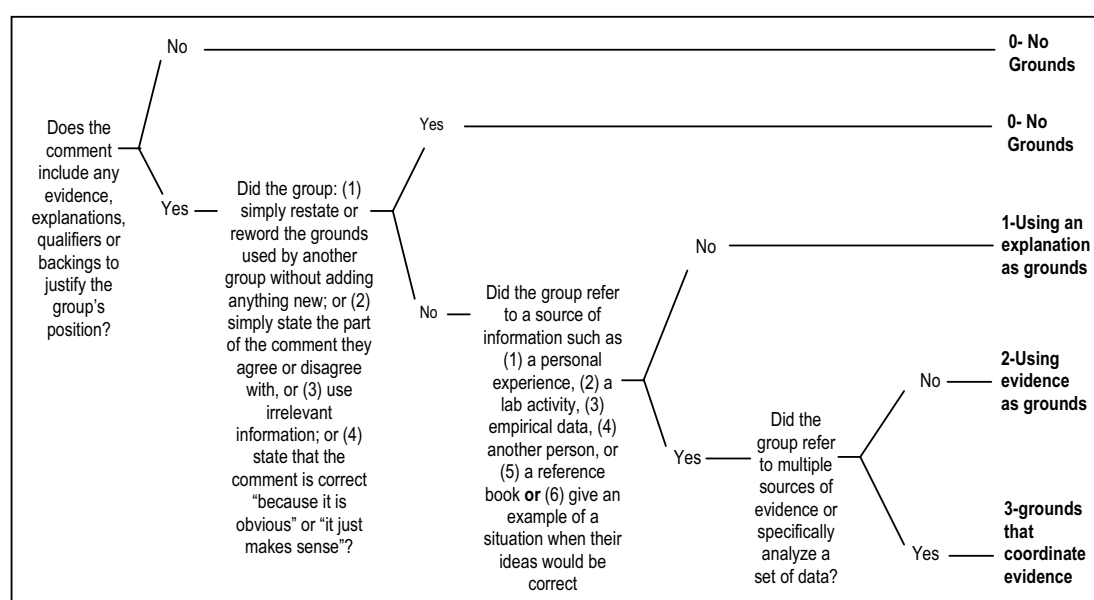


Figure 3. Flow chart for coding the grounds of an individual comment

### Coding the conceptual quality of a comment

Finally, the overall conceptual quality of the comment is rated as either: non-normative (level 0), transitional (level 1), normative (level 2), or nuanced (level 3) after the structure of the groups' comment has been characterized. The coding of conceptual quality involves a coding key of the facets of students' statements similar to Jim Minstrell's facet analysis that was developed for thermodynamics as part of earlier work (Clark, 2000; Clark submitted). In coding a comment, we first determine how many non-normative, transitional, and normative facet are included as part of the entire comment using the facet tables developed through our earlier work (Clark, 2003). These facet tables are not included here due to space restrictions, but may be accessed online through the URL in the reference section. These papers also include further information about coding content facets as non-normative, transitional, normative, and nuanced. After coding the individual facets of a comment, the overall conceptual quality of a comment is determined using the flow chart in Figure 4. The flow chart assigns an overall conceptual quality score based on the frequency of non-normative, transitional, and normative facets found within the entire comment. If the comment does not make sense or the reader can not determine what the authors of a comment are trying to say, the comment is scored as non-normative (0).

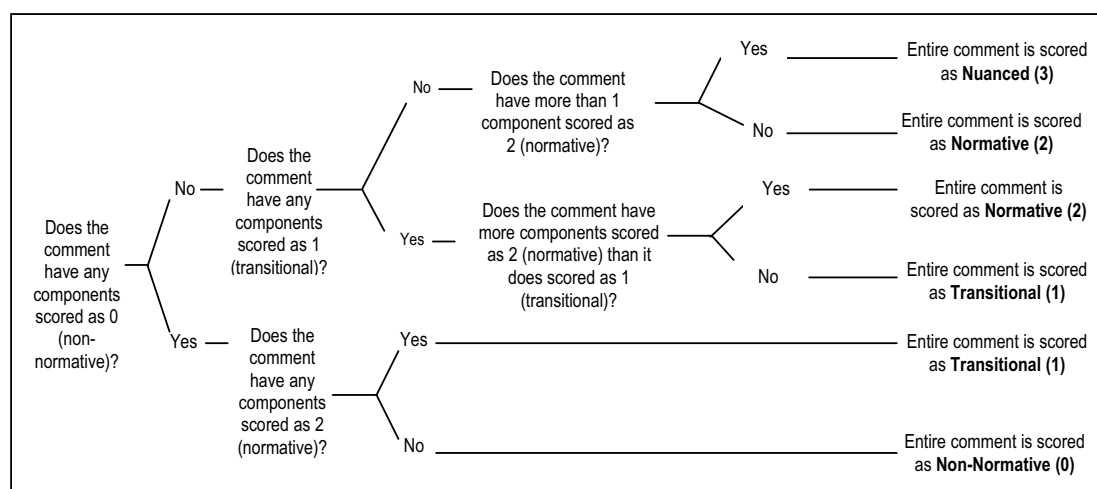


Figure 4. Flow chart for coding the conceptual normativity of an individual comment based on its facets

### Parsing and Scoring Oppositional Episodes

After coding the individual comments, we then code the larger episodes within which the comments occur. In particular, we are interested in oppositional episodes. One challenge involves creating an objective parsing scheme to define episodes. These discussions are threaded and asynchronous. That means that the students may respond to any contribution in their discussion at any time. As is typical in asynchronous threaded forums, responses are placed by the software underneath the parent comment and indented. A fragment of a typical discussion generated from one initial seed claim is outlined in Figure 5. The current study considers the discussion fragment defined by the 2nd level comments (including its parent claim and its children) to be the unit of analysis (i.e., one episode). In Figure 5, there are therefore two episodes defined by 1.1 and 1.2. The 1.1 episode includes 1 and 1.1 only, while the 1.2 episode includes 1, 1.2, and 1.2.1. Each of these episodes is analyzed as a potential oppositional episode. In the Figure 5 example, the episode defined by 1.2 contains opposition while the 1.1 episode does not. Within the actual discussions, a time stamp accompanies each comment to establish the precise time of contribution.

Using the coding schemes for the individual comments, Group 1's comment is a *Claim* with transitional content and no grounds. (Claims created through the principal-maker interface are considered not to include grounds because the student could not add them.) Group 2's comment is *Support* with transitional content and level 2 grounds because they cite the lab results. The episode defined by Comment 1.1 contains only these two comments. This episode contains no opposition and is therefore coded as a non-oppositional episode.

The other episode in the example is defined by Comment 1.2 and includes 1, 1.2, and 1.2.1. It includes opposition and so is coded as an oppositional episode. Group 3's comment is coded as a *Rebuttal Against Thesis* to the "Immediately" part of Group 1's initial claim. Group 3 supports its rebuttal with an explanation about conductivity affecting the rate of temperature change but no evidence (grounds level 1). The conceptual quality of Group 3's comment is coded as being *nuanced* because they claim that the two objects will reach thermal equilibrium but the rate will be influenced by an object's conductivity.

<i>Comment: 1</i> <i>Epistemic: Claim</i> <i>Grounds: Level 0</i> <i>Conceptual Quality: Transitional</i>	<b>Group 1</b> ( <i>This initial statement by a student pair (group 1) is the principle they created to describe their lab data. The principles created then became the seed comments in an online threaded discussion.</i> ) Immediately all objects in the same surround at room temperature become within a few degrees of the same temperature unless an object produces its own heat energy. At this point the objects are within a few degrees even though they may feel different.
<i>Comment: 1.1</i> <i>Epistemic: Support</i> <i>Grounds: Level 2</i> <i>Conceptual Quality: Transitional</i>	<b>Group 2</b> ( <i>response to Group 1</i> ) We agree with this because in the lab we observed that almost all objects were around the same temperature.
<i>Comment: 1.2</i> <i>Epistemic: Rebuttal against claim</i> <i>Grounds: Level 1</i> <i>Conceptual Quality: Nuanced</i>	<b>Group 3</b> ( <i>response to Group 1</i> ) How do you know that the temperature changes immediately? Wouldn't it change at different rates depending on how good a conductor the object is? Couldn't it reach the temperature at a slower or faster rate, although it will eventually reach the same or close to the same temperature of the room?
<i>Comment: 1.21</i> <i>Epistemic: Support</i> <i>Grounds: Level 1</i> <i>Conceptual Quality: Transitional</i>	<b>Group 4</b> ( <i>response to Group 3</i> ) You're right!!! The materials' ability to conduct heat will determine how fast it will heat up. Just like with electricity.

**Figure 5:** Coding of the comments associated with one seed claim

Once individual codes are assigned, overall quality within the oppositional episode can then be analyzed using Table 2 below which we created based on a structural hierarchy developed by Erduran, Osborne, and Simon (Erduran et al., 2004; Simon et al., 2002). After assigning the structural quality of each oppositional episode, our current study then analyzes correlations between the structural quality of the episode in which a comment occurs and the quality of the content and grounds within that comment.

**Table 2:** Assigning overall structural argumentation score to an episode.

Level 5	Argumentation that displays an extended argument that includes multiple rebuttals attacking claims and at least one rebuttal that attacks the grounds used to support a claim
Level 4	Argumentation consists of an extended argument that includes multiple rebuttals that attack claims however there are no rebuttals against the grounds used to support a comment
Level 3	Argumentation has arguments with a series of claims or counter-claims that include grounds with only a single rebuttal
Level 2	Argumentation has arguments with claims or counter-claims with grounds but no rebuttals
Level 1	Argumentation consists of arguments that are simple claim versus counter-claim. There are no grounds or rebuttals included.

## RESULTS AND DISCUSSION

We recently completed the coding of the comments and episodes, and we now present the initial analysis of the data. We are currently employing more sophisticated statistical techniques to study the interconnections between structural quality of the argumentation in an episode and the epistemic operation, grounds, and conceptual quality of the constituent comments in a more nuanced manner, and we will discuss this analysis at the conference.

### The nature of the arguments constructed by students in the PSD environment in terms of context specific content and students' epistemological ideas about argumentation

The comments from the eight discussions are organized by epistemic operation (Figure 6), type of grounds included (Figure 7), and conceptual quality score (Figure 8). The eight discussions involve a total of 334 comments. All 334 comments received an "epistemic operation" code, and 269 of the 334 comments received a code for the type of grounds included and conceptual quality (because certain epistemic types, such as *Organization of Participation*, *Query about Meaning*, and *Off Task*, are not coded for grounds and conceptual quality.) This analysis indicates that students tend to challenge the thesis of a comment (*Rebuttal Against Thesis*: 92) rather than challenge the grounds used by another group (*Rebuttal Against Grounds*: 37). Support of a Comment (69) is the next most common epistemic operation, followed by *Off Task* comments (42) and *Claims* (36). *Query about Meaning* (17) and *Organization of Participation* (7) were less frequent, but played important roles in moving the discussion forward. Out of the 36 initial *Claims*, we do see 10 *Change of Claims*, which suggests that students are willing to consider revising their ideas based on the discussions.

In terms of providing grounds, similar to other research that suggests that students do not usually provide warrants for their claims unless they are challenged (Kelly et al., 1998), students only supported their comments with grounds approximately 51% of the time. When grounds are included as part of a statement, students rely on an unsubstantiated explanation (a causal mechanisms for why a comment is true) rather than including evidence (facts from a source) to support their ideas 47% of the time. In terms of conceptual quality,

student comments are spread fairly evenly in terms of non-normative (29%), transitional (28%), and normative (39%) conceptual quality. Only 4% of the comments are of nuanced conceptual quality.

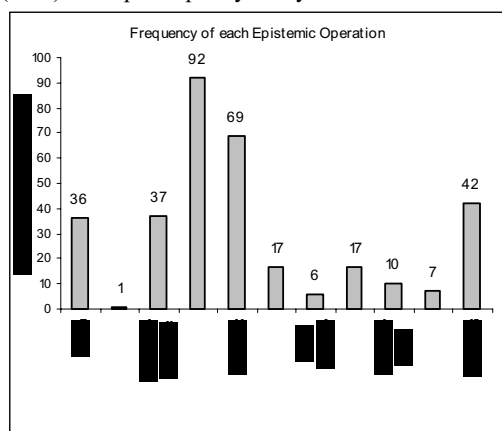


Figure 6: Frequency of comments in terms of epistemic operation

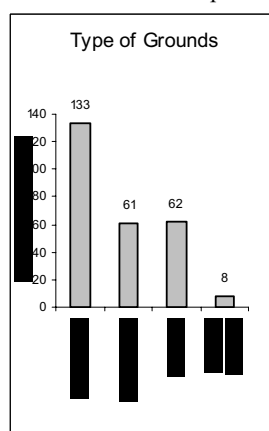


Figure 7: Frequency of the of different types of grounds included in a comment

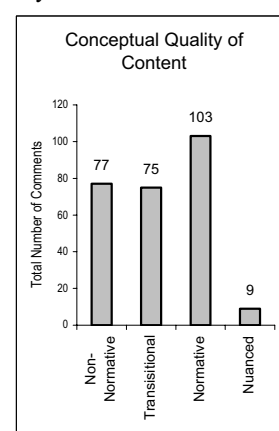


Figure 8: Frequency of comments in terms of conceptual quality

### The relationship between the structural quality of the episode in which a comment occurs and the quality of the comment's grounds and content

The discussions analyzed as part of this study include 126 total episodes involving 334 student comments. Of these episodes, 66 qualify as oppositional episodes and 60 do not. Most non-oppositional episodes tend to be very short (mean number of comments = 2.03) and students do not usually include grounds in order to support their comments (78% of the comments). The conceptual quality of the comments in these episodes tends to be transitional (60% of the comments) or non-normative (21% of the comments) in nature (see Figure 9 and 10). In summary, the non-oppositional episodes tend to be relatively unsophisticated in terms of scientific discourse structures. Students tend to accept what is written in the claim and move onward.

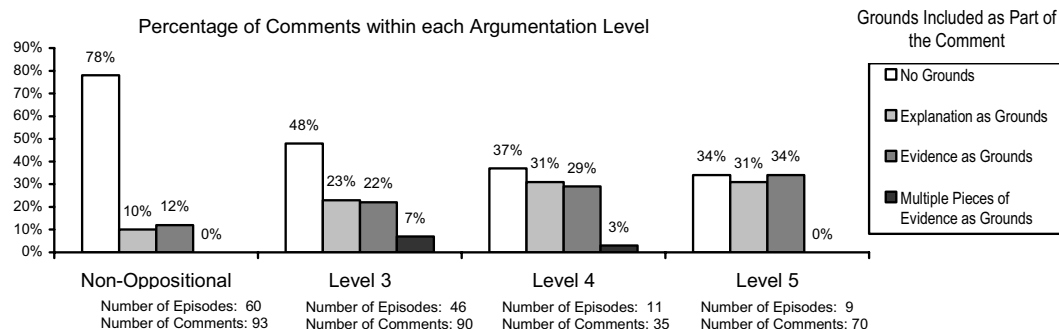


Figure 9: Percentage of the total comments within an argumentation level in terms the type of grounds included

On the other hand, oppositional episodes are longer. For example, the 9 episodes involving multiple sequential rebuttals (level 5) have a mean number of comments of 8.11. This much longer average is heavily weighted by the single longest episode, which spanned 26 comments. Overall, the oppositional episodes include a greater percentage of comments that include grounds (see Figure 9). In level 3 episodes, 53% of the comments include grounds; in the level 4 episodes, 63% of the comments include grounds, and in the level 5 episodes 65% of the comments included grounds. Overall, comments in non-oppositional episodes had a mean grounds score of .38 (SD = .75), comments in the level 3 argumentation episodes had a mean score of .88 (SD = .98), comments in the level 4 argumentation episodes had a mean score of .97 (SD = .89), and comments in level 5 episodes had a mean score of 1.00 (SD = .83). A one way analysis of variance (ANOVA) indicates that these differences in the grounds included by the students are statistically significant,  $F(3) = 7.284$ ,  $p < 0.001$ . These results suggest that in terms of structure, students are able to construct structurally sophisticated arguments.

Within the personally-seeded online discussion forums, students tend to support their statements by including relevant grounds.

In terms of conceptual quality, we also see a difference between oppositional and non-oppositional episodes (Figure 10). In level 3 episodes 42% of the comments coded for conceptual quality were scored as either normative or nuanced. In level 4 episodes the percentage rose to 49%, and in level 5 episodes 61% of the comments were either normative or nuanced. In addition, the percentage of transitional comments differed between non-oppositional episodes (60%), level 3 episodes (22%), level 4 episodes (14%), and level 5 episodes (13%). An additional analysis is underway in order to determine if persuasive students are able to lead other students to more normative understanding (or astray with non-normative grounds).

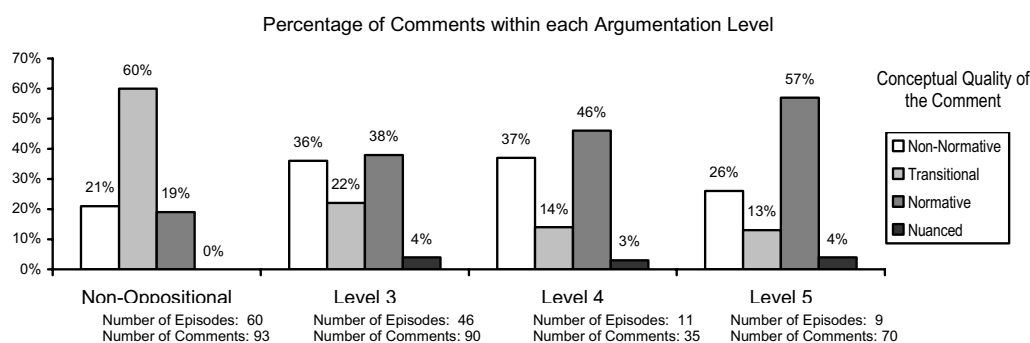


Figure 10: Percentage of the total comments within an argumentation level in terms of conceptual quality

## Appropriateness and Reliability of the Coding Methods Developed and Presented

From a methodological standpoint, the practice of defining episodes based on the second-level comments seems appropriate given the average number of comments and the average depth of comment chains within the episodes in this study. The difference between mean number and mean depth is about 0.1 for non-oppositional episodes and oppositional episodes of level 3 or less. The difference for level 4 arguments is 0.2. These small differences between depth (length of longest chain in episode) and number of comments (total comments in episode) suggest that episodes tend to be linear rather than branched. This linear quality suggests that the current study's parsing scheme seldom combines significantly branched discussions as a single episode. Only the level 5 episodes approach a grey area, where the mean number of comments is 8.11 and the mean depth is 4.33 for the 18 episodes. One of these episodes is unusually large in comparison to the others, containing 26 comments on its own and a depth of 7 with multiple branches. It might be more appropriate to subdivide episodes of this size into multiple episodes representing each of the substantial branches within the large episode. Further analysis of extended episodes will be required to resolve this particular issue. Overall, however, this issue applied to only one episode out of 122 and the parsing method for defining discourse episodes proved appropriate for the vast majority of the episodes.

In order to establish the inter-rater reliability of the coding scheme, the eight discussions were independently coded by the two authors and compared. In spite of the complexity of the proposed coding scheme, inter-rater reliability using this coding system is high. In terms of epistemic operation codes, inter-rater reliability was initially 93%. The largest category of difference between the two coders involved distinguishing off-task comments versus supportive comments. By refining that definition in the coding scheme to include comments that thanked the parent comment's authors for providing support as *Support* rather than *Off Task*, inter-rater reliability climbed to 94%. The remaining 6% of differences were resolved through discussion. This remaining variance between coders seems relatively inevitable but well within an acceptable range.

The initial inter-rater reliability for the coding of grounds was 81%. The largest category of disagreement involved several instances where one author assigned the "explanation" code where the other author assigned the "evidence" code. By refining the evidence definition to include hypothetical examples, the inter-rater reliability climbed to 92%. The next largest category of difference involved one author assigning a "grounds" code where the other author assigned the "no grounds" code. To address the "grounds" versus "no grounds" issue, we revised our definitions to clarify that restating grounds previously included in the episode does not qualify as adding grounds to the argument, this resulted in an increase in inter-rater reliability to 95%. The remaining differences were resolved through discussion. Similar to other studies, we believe that the remaining 5% of variance between coders about the type (or presence) of grounds is acceptable, given the difficulties frequently cited in the literature in terms of the problematic nature of coding grounds. The coding scheme for the individual comments therefore appears to be highly reliable for trained coders in spite of its complexity. The coding of the individual comments draws on the work of many theorists. It is our hope that this



coding scheme will continue to evolve and provide a tool for other researchers interested in the interplay of epistemic operation, grounds, and content.

In terms of scoring of the oppositional episodes, we developed our hierarchy of argumentation structure based on work by Erduran, Osborne, and Simon, but our coding scheme and hierarchy diverges from theirs in some important respects. From their perspective, only arguments that rebut the grounds of another person's argument can undermine the beliefs of that individual. In other words, oppositional episodes that do not rebut the grounds have no potential to change the thinking of the participants because the basis of each participant's beliefs rests on the grounds used as justification. This definition of a rebuttal seems appropriate for debates steeped in social values (e.g., the "socio-scientific" debates in Erduran et al.'s curriculum about whether zoos are good or bad). Erduran et al. define socio-scientific debates based on Monk (1997) as debates "examining and reforming social practices in the light of scientific evidence available through the press and other media." In socio-scientific debates, attacking a grounded claim (e.g., "zoos are good because people can see the animals and want to protect them") with a grounded reply (e.g., "zoos are bad because the animals are unhappy") is often a counter claim rather than a rebuttal. The attack presents another perspective but does not disqualify the initial claim and therefore fits with Erduran et al.'s coding definition that only comments that attack grounds can be coded as rebuttals.

Our study focuses on debates that Erduran et al. would term "scientific" that require empirical argumentation concerning the concept of thermal equilibrium. We define two types of rebuttals: (a) *Rebuttal Against Grounds* which attack the grounds supporting the parent comment's claim and (b) *Rebuttal Against Thesis* which directly rebut the claim of the parent comment. This definition is appropriate in an empirical context because grounds can be provided to fully refute the original claim. For example, a claim that "objects stay different temperatures even if you leave them out on the table for a long time because I've felt them and they feel different" can be rebutted by saying that "the objects actually become the same temperature like when we did the lab and the temperatures of the wood table and the bottle of soda both became 23 degrees after a long time" or by saying that "the objects only feel different even though they are the same temperature because they have different thermal conductivities." From our perspective, both the first reply attacking the claim and the second reply attacking the grounds constitute rebuttals of the initial claim that the "objects stay different temperatures."

Because our definition of rebuttals includes attacks on the claim in addition to attacks on the grounds supporting the original claim, however, our version of the coding scheme results in an elevation in the ranking of some of the episodes in comparison to the results of Erduran et al.'s hierarchy. We acknowledge Erduran et al.'s rationale for coding social debates but assert that our definition correctly values the epistemic operation of attacking a portion of the claim directly in this type of debate, particularly when accompanied by appropriate grounds. We have discussed this issue with Jonathon Osborne (2004) of Erduran et al. (2004) in person, but further work will be required to refine the value and quality codings for the valid epistemic moves that students make in argumentation. Regardless, because of this difference in coding definitions, we do not intend for the scores to be directly compared in terms of which curriculum resulted in "higher" or "lower" scores. Rather, the scores can only be compared qualitatively simply to suggest that the personally-seeded discussions result in successful levels of argumentation, particularly in light of the scientific context, which Simon, Erduran, and Osborne (2003) found to be more challenging for students than socio-scientific contexts.

## CONCLUSIONS

This paper continues the discussion about creating and assessing effective environments to support science inquiry and argumentation. While in-class inquiry discourse typically involves only a small percentage of the students and marginalizes many of the other class members, text-based environments offer the possibility of supporting a much broader range of students (Hsi & Hoadley, 1997). Text-based collaborative environments offer a natural choice because they allow students to participate directly in the linguistic medium of scientific discourse while engaging in inquiry and argumentation. If discourse is important to science, then the opportunity to interact with the actual medium and process of scientific discourse is exceptionally valuable. The results of this study suggest that carefully structured online environments can effectively scaffold student participation in this scientific discourse. The inter-rater reliability results suggest that the proposed coding scheme is also robust in spite of the complex detail with which it analyzes students' participation.

## REFERENCES

- Bell, P. (1997). *Using argument representations to make thinking visible for individuals in groups*. Paper presented at the Computer Supported Clever to Learning Conference '97, Toronto, Canada: University of Toronto.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797-818.
- Clark, D. B. (2000). *Scaffolding knowledge integration through curricular depth*. Unpublished doctoral dissertation, University of California, Berkeley, CA.

- Clark, D. B. (2001). New representations of student knowledge integration in CLP: Theories or repertoires of ideas? Paper presented at the AERA, Seattle, WA.
- Clark, D. B. (2003). *Analyzing student knowledge integration: Theories or pieces?* Paper presented at the National Association of Research in Science Teaching Conference, Philadelphia, PA.
- Collison, G., Elbaum, B., Haavind, S., & Tinker, R. (2000). *Facilitating online learning: Effective strategies for moderators*. Madison, WI: Atwood Publishing.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-213.
- Edelson, D. C., Pea, R. D., & Gomez, L. (1996). The collaboratory notebook: Support for collaborative inquiry. *Communications of the ACM*, 39, 32-33.
- Erduran, S., Osborne, J., & Simon, S. (2004). The role of argument in developing scientific literacy. In K. Boersma, O. deJong, H. Eijkelhof & M. Goedhart (Eds.), *Research and the quality of science education*. Dordrecht: Kluwer Academic Publishers.
- Forman, E. A., Larreamendy-Joerns, J., Stein, M., K., & Brown, C., A. (1998). "you're going to want to find out which and prove it": Collective argumentation in a mathematics classroom. *Learning and Instruction*, 8(6), 527-548.
- Guzdial, M., Turns, J., Rappin, N., & Carlson, D. (1995). Collaborative support for learning in complex domains. In J. L. Schanse & E. L. Cunnius (Eds.), *Proceedings of CSCL '95: Computer supported collaborative learning* (pp. 157-160). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hoadley, C. M., Hsi, S., & Berman, B. P. (1995). The multimedia forum kiosk and speakeasy. In *Proceedings of ACM Multimedia '95* (pp. 363-364). New York: ACM Press.
- Hsi, S., & Hoadley, C. M. (1997). Productive discussion in science: Gender equity through electronic discourse. *Journal of Science Education and Technology*, 6(1), 23-36.
- Jimenez-Alexandre, M., Rodriguez, M., & Duschl, R. A. (2000). 'doing the lesson' or 'doing science': Argument in high school genetics. *Science Education*, 84(6), 757-792.
- Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, 20(7), 849-871.
- Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86, 314-342.
- Lewis, E. L. (1996). Conceptual change among middle school students studying elementary thermodynamics. *Journal of Science Education and Technology*, 5(1), 3-31.
- Lewis, E. L., Stern, J., & Linn, M. C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology*, 33(1), 45-58.
- Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Monk, M., & Osborne, J. (1997). Placing the history and philosophy of science in the curriculum: A model for the development of pedagogy. *Science Education*, 81(4), 405-424.
- Osborne, J., & Clark, D., B. (2004). NSF center for learning and teaching principle investigator conference (pp. Personal Conversation). Washington, DC.
- Resnick, L., B., Salmon, M., Zeitz, C., M., Wathen, S., H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11(3 & 4), 347-364.
- Salmon, G. (2000). E-moderating: The key to teaching and learning online. London; Sterling, VA: Kogan Page.
- Sandoval, W., A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5-51.
- Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom into world 3. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 201-228). Cambridge, MA: MIT Press.
- Schwarz, B., B., Neuman, Y., Gil, J., & Iiya, M. (2003). Construction of collective and individual knowledge in argumentative activity. *Journal of the Learning Sciences*, 12(2), 219-256.
- Simon, S., Erduran, S., & Osborne, J. (2002). *Enhancing the quality of argumentation in school science*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, April 7-10, New Orleans, USA.
- 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 and New York: Routledge Falmer.
- Tabak, I., Smith, B. K., Sandoval, W. A., & Reiser, B. J. (1996). Combining general and domain-specific strategic support for biological inquiry. In C. Frasson, G. Gauthier & A. Lesgold (Eds.), *Proceedings of the third international conference on intelligent tutoring systems (its '96)*. Montreal, New York: Springer-Verlag.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- 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.