Assessing Dialogic Argumentation in Online Environments to Relate Structure, Grounds, and Conceptual Quality

Douglas B. Clark,[1] Victor Sampson[2]

[1]College of Education, Payne 203F, Arizona State University, Tempe, Arizona 85287-0911

[2]The Florida State University, Department of Middle and Secondary Education, 209 MCH, Tallahassee, Florida 32306-4490

Received 15 September 2006; Accepted 3 April 2007

Abstract: The national science standards, along with prominent researchers, call for increased focus on scientific argumentation in the classroom. Over the past decade, researchers have developed sophisticated online science learning environments to support these opportunities for scientific argumentation. Assessing the quality of dialogic argumentation, however, has proven challenging. Existing analytic frameworks assess dialogic argumentation in terms of the nature of students' discourse, formal argumentation structure, interactions, and epistemic forms of reasoning. Few frameworks, however, connect these assessments to conceptual quality. We present an analytic framework for assessing argumentation in online science learning environments that relates levels of opposition with discourse moves, use of grounds, and conceptual quality. We then apply the proposed framework to students' dialogic argumentation within a representative online science learning environment to investigate the framework's potential affordances as well as to assess issues of reliability and appropriateness. The results suggest that the framework offers significant affordances and that it also offers high interrater reliability for trained coders. The applicability of the framework for offline contexts and future extensions of the framework are discussed in light of these results. © 2007 Wiley Periodicals, Inc. J Res Sci Teach 45: 293–321, 2008
Keywords: physical science; cooperative grouping; technology education/software design

How can we assess and characterize the nature of students' argumentation within online science learning environments? In particular, is there a way to examine the relationships that exist between levels of opposition and the types of comments students make, their use of grounds, and the conceptual quality of their comments?

Argumentation is considered a key component of science education by the national science standards (American Association for the Advancement of Science, 1993; National Research

Contract grant sponsor: United States National Science Foundation; Contract grant number: REC-0334199 (TELS: The Educational Accelerator: Technology-Enhanced Learning in Science).
Correspondence to: D.B. Clark; E-mail: dbc@asu.edu
DOI 10.1002/tea.20216
Published online 5 December 2007 in Wiley InterScience (www.interscience.wiley.com).

Council, 1996) and by many prominent researchers in science education (Driver, Newton, & Osborne, 2000; Duschl & Osborne, 2002; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000; Kuhn, 1993; Siegel, 1989). Research suggests that online learning environments can provide excellent support for argumentation in the classroom (Andriessen, Erkens, Van de Laak, Peters, & Coirier, 2003; Bell & Linn, 2000; Clark & Sampson, 2005; deVries, Lund, & Baker, 2002; Hoadley & Linn, 2000; Sandoval & Reiser, 2004; Schwarz, Neuman, Gil, & Ilya, 2003; Veerman, 2003). Measuring the quality of students' scientific argumentation online and offline, however, proves challenging (Andriessen, Baker, & Suthers, 2003; Erduran, Osborne, & Simon, 2004; Kelly, Druker, & Chen, 1998; Sandoval & Millwood, 2005).

In this study, we present an analytic framework for coding students' dialogic scientific argumentation in the asynchronous threaded forums of online science learning environments. The framework relates the level of opposition found within a discourse episode and the types of comments students make, the grounds quality included in those comments, and the conceptual quality of their ideas. More specifically, our goals include: (1) presenting the analytic framework; (2) applying the analytic framework to argumentative discussions occurring within the asynchronous forums of a representative online science learning environment; (3) considering the implications of the findings for the environment; (4) investigating the reliability, affordances, and constraints of the proposed analytic framework through this process; and (5) discussing future directions, including the adaptation of the framework to standard classroom contexts.

Background

Online learning environments designed to engage and support students in dialogic argumentation provide excellent opportunities for students to productively propose, support, evaluate, critique, and refine ideas. Over the last decade, a number of sophisticated environments have been developed to support students engaging in this type of knowledge-building discourse. Examples, among others, include *CONNECT* (e.g., deVries et al., 2002), TC3 (e.g., Erkens, Kanselaar, Prangsma, & Jaspers, 2003), *DUNES* (e.g., Schwarz & Glassner, in press), the Virtual Collaborative Research Institute (e.g., Janssen, Erkens, Jaspers, & Kanselaar, 2006), *ArgueGraph* (e.g., Jermann & Dillenbourg, 2003), and the *personally-seeded discussions* within the Web-based Inquiry Science Environment (e.g., Clark & Sampson, in press).

These online learning environments provide a broad range of specific instructional functionalities to promote dialogic argumentation and facilitate active learning potentially beyond what can be achieved in traditional learning environments (Fabos & Young, 1999; Fischer, 2001; Marttunen & Laurinen, 2001; Pea, 1994; Roschelle & Pea, 1999; Schellens & Valcke, 2006). These environments integrate multiple categories of functionalities to support argumentation, including *asynchronous and synchronous collaborative communication interfaces* (e.g., deVries et al., 2002; Joiner & Jones, 2003; Marttunen & Laurinen, 2001; Schellens & Valcke, 2006), the *co-creation and sharing of intellectual artifacts* (e.g., Erkens et al., 2003; Kirschner, Buckingham, Shum, & Carr, 2003; Schwarz & Glassner, 2003), *enriched access to information*; Kolodner, Schwarz, Barkai, Levy-Neumand, Tcherni, & Turbovsk, 1997; Oestermeier & Hesse, 2000), *collaboration scripting functionalities* (e.g., Clark & Sampson, 2006a; Jermann & Dillenbourg, 2003; Stegmann, Weinberger, & Fischer, 2005), and *awareness heightening tools* (Erkens & Janssen, 2006; Hesse, 2007; Jermann, Soller, & Muehlenbrock, 2001). It is important to note that not all online environments designed to foster dialogic argumentation incorporate all of these features. Rather, each environment integrates subsets of these features (and potentially others) to support the designers' specific theoretical perspectives on argumentation and pedagogical goals. Thus, in practice, these online learning environments engage complex interrelationships of these functionalities. Clark, Sampson, Weinberger, and Erkens (in press) provide an extended discussion of these functionalities.

The complexity of issues just discussed gives rise to complex and diverse assessment needs for researchers examining dialogic argumentation in these contexts. To date, researchers have developed a broad range of methods to assess the nature of dialogic argumentation within online and offline contexts that reflect various perspectives on argumentation, pedagogical goals, and curricular structures. Robust analytic frameworks exist that examine formal argumentation structure (e.g., Erduran, Simon, & Osborne, 2004; Osborne, Erduran, & Simon, 2004), the nature and function of contributions within the dialog (e.g., Baker, Andriessen, Lund, van Amelsvoort, & Quignard, in press; deVries et al., 2002), the epistemic nature of reasoning (e.g., Duschl, in press; Erkens et al., 2003; Jimenez-Aleixandre et al., 2000), and the patterns and trajectories of participant interaction (e.g., Baker, 2003; Hogan, Nastasi, & Pressley, 2000; Leitão, 2000; Weinberger & Fischer, 2006). Clark, Sampson, and Weinberger (in press) outline and catalog a broad range of these methods and address the potential benefits of these various analytic foci for the assessment of dialogic argumentation in online settings.

Although several strong analytic frameworks focus on these various aspects of argumentation, fewer analytic frameworks focus on the conceptual quality of students' contributions in relationship to these other aspects of dialogic argumentation. Essentially, although there are promising methods for assessing the conceptual quality of a rhetorical argument produced by students to support or refute an idea (e.g., Sandoval & Millwood, 2005), options for dialogic argumentation remain limited. Kuhn and Udell's (2003) and Zohar and Nemet's (2002) frameworks represent two of the more prominent approaches for dialogic argumentation that incorporate a focus on conceptual quality (although they apply their framework in offline contexts). Kuhn and Udell's study (2003) compares the impact on subjects' final arguments of engaging in dialogic argumentation with a partner who holds an opposing view. Zohar and Nemet's (2002) study compares the impact of integrating explicit instruction about general reasoning patterns into the teaching of specific science content (and includes dialogic argumentation within the intervention, although it is not the focus). Both of these frameworks, however, measure the conceptual quality of students' arguments before and after their interventions rather than assessing the conceptual quality of the students' contributions within the dialog. Our proposed framework provides an alternative approach and alternative affordances by focusing on the conceptual quality of comments within the dialog and focusing more specifically on a graded spectrum of scientific normativity.

Our prior work (Clark & Sampson, in press) suggested that students' arguments often incorporate inaccurate scientific ideas even though their argument is relatively sophisticated from a structural perspective. 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 a structural perspective, a rebuttal like this is excellent in terms of the way it incorporates data and warrants. The student in this example, however, confuses electrical and thermal conductivity. Students also tend to distort data to support their idiosyncratic ideas (Carey, Evans, Honda, Jay, & Unger, 1989; Chinn & Brewer, 1998; Clark & Sampson, 2005, 2006a; Schauble, Glaser, Duschl, Schulze, & John, 1995). Our proposed framework therefore parses, codes, and analyzes student argumentation based on discourse moves, use of evidence as grounds, and conceptual quality. In this way, the framework not only examines the structural components of argument (e.g., claims, grounds, counterclaims, and rebuttals) that students use as they propose, support, evaluate, and critique ideas during a discussion, but also the kinds of grounds that students use and the conceptual quality of their ideas.

Proposed Analytic Framework

The proposed framework analyzes dialogic scientific argumentation within asynchronous threaded online forums. As discussed previously, the framework focuses on the relationships between levels of opposition found within a discourse episode (e.g., no opposition at all or multiple challenges) and the types of comments students make, the grounds quality included in those comments, and the conceptual quality of their ideas. By focusing on the relationships between these aspects of argumentation, the framework offers researchers a specific analytic tool to examine possible connections between argumentation and subject matter learning. We are therefore not claiming superiority of the framework in comparison to the existing excellent frameworks upon which it builds. Rather, we are simply claiming that the framework contributes specific functionality that is currently underrepresented in the overall set of tools available to researchers in this field. The discussion of the framework first outlines our theoretical perspectives on dialogic argumentation, then focuses on the coding of individual comments, then discusses the parsing and coding of larger discourse episodes, and finally outlines the analytic approach used to investigate the relationships between discourse episodes and constituent comments.

Our Theoretical Perspective on Dialogic Argumentation

To guide our work we have adopted a view of dialogic argumentation as a process whereby ‘‘different perspectives are being examined and the purpose is to reach agreement on acceptable claims or course of actions’’ (Driver et al., 2000, p. 291). This interpretation defines dialogic argumentation as a social and collaborative process that is used ‘‘to solve problems and advance knowledge’’ (Duschl & Osborne, 2002, p. 41) rather than as an effort to ‘‘justify or refute a particular standpoint’’ (van Eemeren, Grootendorst, & Henkemans, 2002, p. 38) or as the articulation of informal reasoning (e.g., Perkins, Farady, & Bushy, 1991; Sadler, 2004; Zohar & Nemet, 2002). This view of dialogic argumentation stresses collaboration over competition and suggests that activities that promote dialogic argumentation can provide a context whereby individuals are able to use each others' ideas to construct and negotiate a shared understanding of a particular phenomenon in light of past experiences and new information (Abell, Anderson, & Chezem, 2000; Andriessen, Baker, & Suthers, 2003; Boulter & Gilbert, 1995; deVries et al., 2002; Veerman, 2003).

Given this theoretical perspective, we have conceptualized dialogic argumentation as a process of proposing, supporting, evaluating, and refining ideas in an effort to make sense of a complex or ill-defined problem. Our efforts to support and promote argumentation in science education have therefore focused on the development of an online learning environment that enables students to generate competing explanations for a given phenomenon, and then provides them with an opportunity to examine, discuss, and evaluate these explanations using available evidence. To support students as they engage in this complex task, we have designed a number of different electronic tools that serve as scaffolds for students as they work. For example, the *personally seeded discussion* was designed to: (a) support students as they synthesize a principle to describe data that they have collected; (b) organize discussion groups of students who have created different principles; and (c) encourage students to critique each other's principles and work toward consensus based on evidence available to them within an asynchronous online discussion forum

(Clark & Sampson, 2006b; in press). To evaluate the effectiveness of this online environment and these electronic tools as a way to foster quality dialogic argumentation, we required an analytical framework that could relate levels of opposition with discourse moves, use of grounds, and the conceptual quality of students' ideas.

Coding the Discourse Moves of Individual Comments

The framework assigns a *discourse move* code to each comment based on the comment's role in the discussion. To avoid ambiguity in terms of references within a comment, the framework codes each comment in relation to the parent comment to which it responds. These codes take into account comments that are typically examined as part of a structural analysis (e.g., claims, counterclaims, rebuttals), meta-organizational comments that that help organize the interaction (which are typically overlooked in a structural analysis), and the occasional off-task interaction.

Our coding of argumentation discourse moves draws heavily on the combined work of Jonathan Osborne, Sibel Erduran, and Shirley Simon (e.g., Erduran et al., 2004; Osborne et al., 2004; Osborne, Simon, & Erduran, 2002; Simon, Erduran, & Osborne, 2002, 2006). Erduran et al. (2004) provided a fully detailed account and analysis of their excellent framework. We extend the list of argumentation discourse codes from Erduran et al. (2004) with meta-organizational and off-task discourse codes that are similar in function and categories to those developed by deVries et al. (2002) and Baker et al. (in press). In addition to extending our discourse codes beyond argumentative moves, our approach diverges from Erduran et al.'s (2004) approach in terms of our addition of a second category of rebuttals. Erduran et al. (2004) defined rebuttals as "challenging the validity of the evidence and justification offered" (p. 921). We incorporate this definition for our *rebuttal against grounds* code. Our added second category of rebuttals (*rebuttal against thesis*) focuses on challenges to the validity of the thesis (or a specific part of the thesis) of a claim or rebuttal.

This expanded definition of rebuttals represents what we perceive as epistemological differences in the argumentation contexts that we are studying in comparison to the focus outlined by Erduran et al. (2004). 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. We agree that this definition of rebuttal seems appropriate when the object of the discussion is a "socio-scientific issue" (Osborne et al., 2002, p. 3) steeped in personal values (e.g., genetic engineering, reproductive technologies, diet choices, or food safety). When the purpose of engaging in argumentation is to reach consensus about a socio-scientific issue (e.g., the socio-scientific issues in their work where students argue about whether zoos are "good" or "bad"), 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 considered a counterclaim rather than a rebuttal. The attack presents another perspective but does not disqualify the initial claim and therefore fits with the Erduran et al. coding definition that only comments that attack grounds can be coded as rebuttals.

Our study focuses on the ways students engage in argumentation when the object of the discussion involves a more "scientific issue" (Osborne et al., 2002, p. 3). In this study, the students engage in argumentation to make sense of a complex or ill-defined problem related to the concept of thermal equilibrium. We propose that these expanded conceptions of rebuttals are appropriate in this type of context because a rebuttal can provide grounds to fully refute the original claim directly without attacking the grounds (and therefore is distinct from a simple counterclaim that merely proposes a different interpretation of the evidence). 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 like the metal spoon and a wood spoon that are the same temperature in the lab.'' The first reply directly rebuts the claim and the second rebuts the grounds. Both constitute epistemologically valid rebuttals (rather than simple counterclaims) of the initial claim that ''objects remain different temperatures'' from our perspective for these ''scientific'' debates. Our framework can be adapted to incorporate Erduran et al.'s rebuttal definition for contexts better suited to that definition, however, as discussed in greater detail in the subsequent section ''Scoring the Level of Opposition Found Within Discourse Episodes.''

Our full list of *discourse move* comment codes is outlined in Table 1. Table 2 provides an example of how the *discourse move* of each comment is coded. Comment #1 is coded as a *claim* because it was created using the principal-maker interface of that particular learning environment. Pair B's comment (#1.1) is coded as a *rebuttal against thesis* to the ''immediately'' part of Pair A's initial claim. Pair C's comment (#1.11) is classified as s*upport* because they voice agreement with Pair B's rebuttal and then add additional grounds in defense of this viewpoint. Finally, comment #1.2 is also coded as *support* because Pair D indicates that they concur with Pair A's initial seed claim.

Coding the Grounds of a Comment

Rather than focusing on the type of epistemic reasoning employed by the student (e.g., Duschl, in press; Erkens et al., 2003; Jimenez-Aleixandre et al., 2000), the framework focuses more specifically on the level of grounds that students include to warrant their claim with a focus on their coordination of evidence with claims. The framework incorporates Osborne et al.'s (2004) strategy of collapsing Toulmin's (1958) data, warrants, and backings into a single category called ''grounds.'' Rather than simply identifying the presence or absence of grounds, the framework classifies a comment as having no grounds (*grounds quality level 0*), including only an explanation without evidence as grounds (*grounds quality level 1*), using evidence as grounds (*grounds quality level 2*), and including evidence and an explanation or coordinating multiple pieces of evidence or as grounds (*grounds quality level 3*). This approach provides an ordinal scale that facilitates statistical analysis. We developed a series of binary decisions (represented in the Fig. 1 flowchart) to increase reliability in the coding process. Whereas all comments receive a discourse move code, not all comments receive a grounds quality and conceptual quality code because comments with certain discourse move codes, such as ''organization of participation,'' ''query about meaning,'' and ''off-task,'' are not eligible for grounds quality or conceptual quality codes in this system.

Table 2 illustrates how the individual comments are coded in terms of *grounds quality* in our discourse example. Comment #1 is coded as *grounds quality level 0* (no grounds) because claims created through the principal-maker interface are considered not to include grounds due to the fact that the student could not add them. Pair B (comment #1.1) supports their rebuttal with a plausible explanation about conductivity affecting the rate of temperature change, but fails to substantiate their explanation by citing data or a source of information such as a lab activity or a reference book so their comment is coded as *grounds quality level 1* (explanation without evidence). Comment #1.11 is also classified as consisting of an explanation without evidence (*grounds quality level 1*) because Pair C's explanation for why objects reach thermal equilibrium at different rates is an inference based on their experiences. Finally, Pair D's comment (#1.2) is coded as *grounds quality*

Table 1
Coding scheme for the discourse move of individual comments

Discourse Move	Definition
Claim	The seed-comment principle or an assertion made by a pair of students.
Counterclaim	An assertion made by a pair of students that is different from (and does not attack) the seed claim or parent comment made by another pair of students. 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.
Change of Claim	A comment made by a pair of students indicating 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 pair of students.
Rebuttal Against Grounds	An attack on, or disagreement with, the grounds (evidence, explanations, qualifiers, or backing) used by another pair of students to support or justify their comment.
Rebuttal Against Thesis	An attack on or disagreement with the thesis (or a specific part of the thesis) of another pair of students' comment (claim or rebuttal) that does not attack the grounds.
Clarification in response to a Rebuttal	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 pair of students.
Support of a Comment	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 on the comment.
Query about Meaning	A comment that asks for clarification of an earlier comment (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.
Clarification of Meaning	A comment made by a pair 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 than supporting the accuracy of a statement.
Organization of Participants	A comment that: (1) reminds other participants to participate; (2) asks others for feedback; (3) has a meta-organizational aspect (e.g., "Do we all agree?"); or (4) attempts to change the way someone else in the discussion is participating.
Off-task	Comments that are not about the topic (e.g., "Nice haircut, John!").

level 2 (evidence as grounds) because they cite their lab results when they indicate that they support Pair A's claim.

Coding the Conceptual Quality of a Comment

Finally, the *conceptual quality* of the comment is rated as either non-normative (*conceptual quality level 0*), transitional (*conceptual quality level 1*), normative (*conceptual quality level 2*), or

Table 2
Discourse example outlining the coding of the comments associated with one seed claim

Comment Number	Comment	Coding
1. *Seed-Comment*	Pair A: Immediately all objects in the same surroundings 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.	*Discourse move*: Claim *Grounds*: None *Conceptual*: Transitional
1.1. *Response to Pair A*	Pair B: 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?	*Discourse move*: Rebuttal against Thesis *Grounds*: Explanation only *Conceptual*: Nuanced
1.11. *Response to Pair B*	Pair C: You're right!!! The materials' ability to conduct heat will determine how fast it will heat up. Insulators heat up quicker than conductors.	*Discourse move*: Support *Grounds*: Explanation only *Conceptual*: Transitional
1.2. *Response to Pair A*	Pair D: We agree with this because in the lab we observed that almost all objects were around the same temperature.	*Discourse move*: Support *Grounds*: Evidence *Conceptual*: Transitional

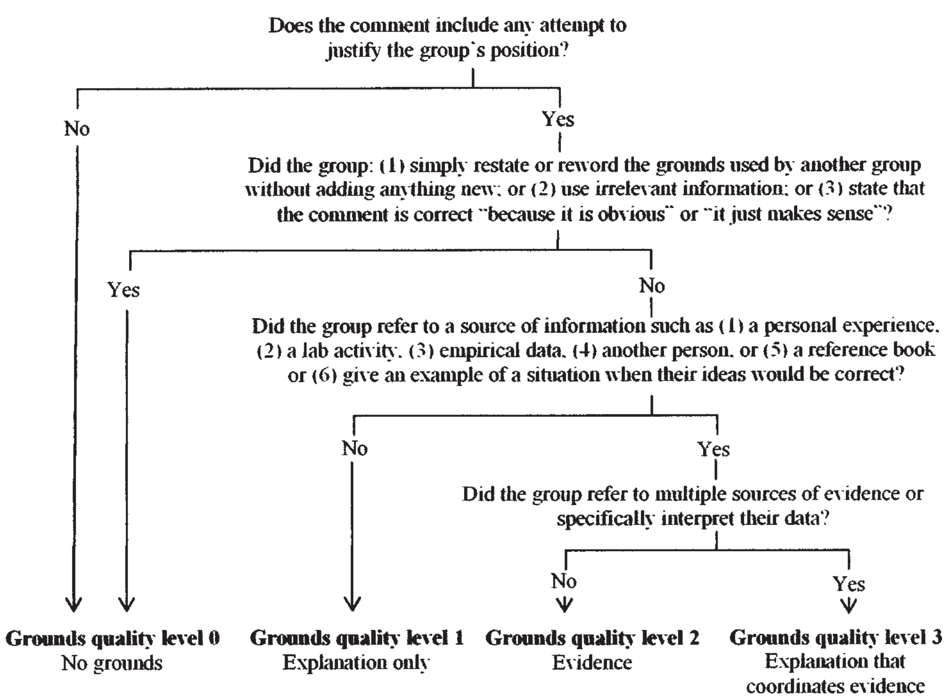

Figure 1. Flowchart for coding the grounds of an individual comment.

nuanced (*conceptual quality level 3*). In coding a comment, the framework first determines how many non-normative, transitional, and normative facets are included as part of the entire comment using conceptual facet tables (see Table 3). This approach builds on extensive prior conceptual change work measuring the longitudinal evolution of students' conceptual ecologies (Clark, 2000, 2001; Clark & Linn, 2003). After coding the individual facets of a comment, the overall *conceptual quality* of a comment is determined through the series of binary decisions represented

Table 3
Example facets for coding conceptual quality of comment

Non-normative facets
Metal objects are above/below ambient temperatures by a great difference
All objects are (end up) way above or way below ambient temperature
Objects have a "natural" temperature
Wool makes things warm
Cold energy travels from cool objects to warm object which cools them down
Thermal equilibrium does not happen to all objects/same room=different temperatures
Objects will take in heat until they "fill up," then they stop heating up
Circulating air within an object moves heat through a medium
Hot things can give off heat energy without cooling (and reverse)
Size/thickness affects final temperature
Objects keep getting hotter/colder past equilibrium temperature the longer they are in a hot/cold place
Objects will "overshoot" equilibrium temperatures as they warm up or cool down
Unless something is actively keeping an object cold, it will warm up
Conductors attract heat/take in a lot of heat/always take in heat
Materials with holes (porous/density) allow heat/cold energy to pass through (barrier model for insulation)
Good conductors/insulators keep heat on the surface versus keeping heat/cold inside
Thickness of a material alone determines insulating/conducting properties
Insulators store heat or cold energy and/or release it slowly
Wool warms things up/good insulators warm things up or cool them down
Metals insulate because they reflect heat away
Metals keep things cold because they feel cold
Metal conducts heat away from cold objects to keep them cold/metal gets cold from object and then keeps object cold
Only cold/hot things can keep objects cold/hot
Conductors heat or cool slowly/insulators heat or cool quickly
Wood/Styrofoam/plastic/glass only take in a little heat
Material that keep things cold cannot also keep things hot (and reverse)
Heat or cold energy cannot travel through metal
The conductivity of an object influences the object's *ability* to reach equilibrium
Transitional facets
All objects in the same room will reach close temperature (but not the same) as surroundings
Wood/wool/plastic/glass/metal will not reach equilibrium because they need more time
Insulators "block, trap, or allow a small amount of" heat or cold (act like barriers)
Heat comes in and out of metal quickly at a constant rate
Metals "adjust" temperature more quickly (without making a connection to conductivity)
Wood/wool/plastic/glass change temperature slower then metal (without making a connection to conductivity)
Styrofoam/wool is good for keeping cold objects cold (without making a connection to conductivity)
Metal is not good for keeping cold objects cold (without making a connection to conductivity)
Normative facets
Objects in the same room become the same temperature
If more heat flows into an object than out of it, its temperature rises (or reverse)
Heat energy flows from hot to cold objects

(*Continued*)

Table 3
(*Continued*)

Heat flows until objects reach the same temperature
Objects do not go beyond equilibrium temperatures
Wood/wool/Styrofoam/plastic/metal objects reach temperature of surroundings
Same room/same temperature unless another heat source or it produces it own heat
Size and thickness do not affect final temperature
Metal conduct heat well
Wood/wool/Styrofoam/plastic are insulators (do not conduct heat well)
Conductors conduct heat energy quickly/insulators conduct heat energy slowly
Good conductors are poor insulators (and reverse)
Objects that keep cold things cold also keep hot things hot
Conductors heat up faster, insulators heat up slower
Rate of reaching equilibrium is dependent on conductivity
Appropriate connection of thickness/size to rate of heating/cooling

Note. Although there are four conceptual quality codes for comments (*non-normative, transitional, normative, nuanced*), there are not "nuanced" facets. A nuanced comment is defined in terms of the presence of multiple normative facets and the absence of non-normative and transitional facets. See Figure 2 for more details.

in the Figure 2 flowchart. The flowchart assigns an overall *conceptual quality* score based on the frequency of non-normative, transitional, and normative facets found within the entire comment. If the coder cannot determine what the students are trying to say in a comment, the comment is scored as non-normative (*conceptual quality level 0*).

As discussed, this approach offers alternative affordances to the approaches of Kuhn and Udell (2003) and Zohar and Nemet (2002), who assessed conceptual quality in terms of pretests and posttests that bookended their interventions. Whereas the approaches of Kuhn and Udell (2003) and Zohar and Nemet (2002) offer affordances for determining trajectories of change in terms of conceptual quality, our proposed approach offers affordances in terms of analyzing connections between specific discourse moves, grounds use, and conceptual quality within the

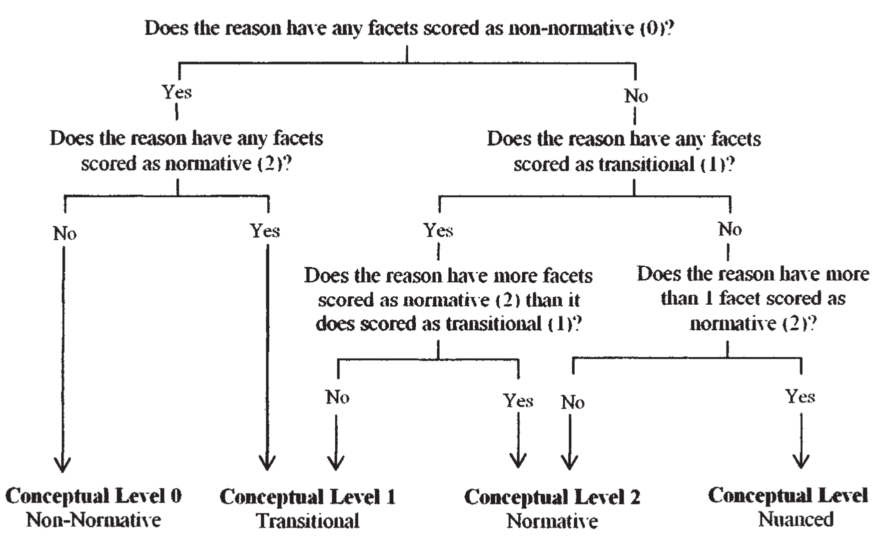

Figure 2. Flowchart for coding the conceptual normativity of an individual comment based on its facets.

dialog. In terms of similarities, all three of the approaches implement parsimonious ordinal scales of conceptual quality to facilitate reliability and statistical comparison.

Examples of how comments are coded for conceptual quality are provided in Table 2. Pair A's initial claim is coded as *conceptual quality level 1* (transitional content) because the comment contains more transitional facets (''immediately all objects in the same surroundings at room temperature become within a few degrees of the same temperature'' and ''objects are within a few degrees even though they may feel different'') than normative facets (''unless an object produces its own heat energy''). The conceptual quality of Pair B's comment (#1.1) is coded as *conceptual quality level 3* (nuanced content) because they claim that the two objects will reach thermal equilibrium and that the rates will be influenced by the objects' conductivities. Pair C's comment (#1.11) is classified as *conceptual quality level 1* (transitional content) because they claim that different materials reach thermal equilibrium at different rates (normative) but they also indicate that ''insulators heat up quicker than conductors'' (non-normative). Finally, Pair D's comment (#1.2) is classified as *conceptual quality level 1* (transitional content) because they voice agreement with a transitional claim and offer further support for the idea that ''in the same surroundings objects become within a few degrees of the same temperature.''

Parsing the Discussions into Discourse Episodes

After coding the individual comments, the framework then codes the larger episodes of discourse within which the comments occur. The framework parses the discussions into discourse episodes (*episode* for short) because we are interested in the relationship between the level of opposition or conflict that takes place within an episode and the nature of the individual constituent comments.

Before we can discuss the episode parsing scheme, we must first clarify the mechanics and terminology of asynchronous threaded online forums. In asynchronous threaded forums, students may respond to any contribution in their discussion at any time. Responses are placed by the software underneath the parent comment, indented, and are accompanied by a time stamp to establish the precise time of contribution. In the discourse example (Table 2), the seed-comment (comment #1) is the first-level comment. Comments #1.1 and #1.2 are second-level comments because they are children-comments of (respond to) the first-level comment. Comment #1.11 is a third-level comment because it is a child-comment of (responds to) a second-level comment. The seed-comments (e.g., comment #1) represent students' opening claims in the discussion.

The framework considers an episode to be defined by each second-level comment (including its parent claim and its children). There are therefore two episodes defined by the second-level comments #1.1 and #1.2 in Table 2. The #1.1 episode includes comments #1, #1.1, and #1.11, whereas the #1.2 episode includes comments #1 and #1.2 only. Once a discussion is parsed into discourse episodes, the *level of opposition* that takes place within each episode is coded using the method outlined in what follows.

Scoring the Level of Opposition Found Within Discourse Episodes

The framework characterizes the amount of conflict or *level of opposition* that takes place within an episode using the hierarchy outlined in Table 4. The framework defines high-quality argumentation (*oppositional level 5*) as discourse that emphasizes the use of multiple rebuttals that challenge the interpretation of a phenomenon and the validity of the grounds that are used to support this interpretation. On the other hand, low-quality argumentation is either

Table 4
The overall quality of the argumentation that takes place within an episode determined using a hierarchy based on opposition

Quality	Characteristics of the Discourse
Level 5	Argumentation involving multiple rebuttals and at least one rebuttal that challenges the grounds used to support a claim
Level 4	Argumentation involving multiple rebuttals that challenge the thesis of a claim but does not include a rebuttal that challenges the grounds used to support a claim
Level 3	Argumentation involving claims or counterclaims with grounds but only a single rebuttal that challenges the thesis of a claim
Level 2	Argumentation involving claims or counterclaims with grounds but no rebuttals
Level 1	Argumentation involving a simple claim versus counterclaim with no grounds or rebuttals
Level 0	Non-oppositional

non-oppositional (*oppositional level 0*) or consists of only claims and counterclaims that do not attempt to challenge the validity of the other participants' interpretation of the phenomenon (*oppositional level 1*). Although we agree with researchers who suggest that non-oppositional discourse is an important part of argumentation because individuals often "build on each other's knowledge" rather than attempting to "convince each other of their own viewpoint" when engaged in argumentation (Andriessen et al., 2003, p. 82), we also believe that "students must learn how to question scientific claims and explore the epistemological underpinnings of scientific knowledge" (Linn & Eylon, 2006, p. 47). We believe that this type of emphasis encourages students to view ideas, opinions, sources of evidence, and reasoning as objects of cognition (Kuhn, 1993; Mason, 2001) and provides a context whereby the epistemic aspects of science make sense (Kuhn & Reiser, 2005, 2006).

This scheme adapts the hierarchy outlined by Erduran et al. (2004) by incorporating our expanded definitions of rebuttals. In the Table 2 example, the episode defined by comment #1.1 contains three comments—a *claim*, a *rebuttal against thesis*, and a *support*. Using the structural hierarchy outlined in Table 4, this episode is therefore coded as *structural quality level 3* (oppositional) because it contains a single rebuttal that challenges the validity of a claim. On the other hand, the episode defined by comment #1.2 is classified as *structural quality level 0* (non-oppositional) because it only includes a *claim* and a comment that offers *support*.

Essentially, we define the overall quality of argumentation by focusing on the presence or absence of both types of rebuttals in our framework rather than classifying *structural quality* solely on the presence of rebuttals challenging the validity of data and warrants. From our perspective, the presence of either type of rebuttal serves as a significant indicator of *structural quality*, because either type of rebuttal can be used to force the author of a claim to evaluate the validity of their argument. Because our definition of rebuttals includes attacks on the claim as well as attacks on the grounds supporting the original claim, our version of the coding scheme results in an elevation in the ranking of some of the episodes in comparison to the framework outlined by Erduran et al. (2004). We acknowledge their rationale for coding social debates but our expanded pair of rebuttal definitions can offer valuable insights into the nature of student discourse in scientific debates. This is obviously a philosophical and epistemological valuation dependent upon the context being studied.

As a side note, researchers who believe that their context more closely aligns with the definitions of Erduran et al. (2004) could still employ their hierarchy and argumentation codes while still taking advantage of our additional meta-organizational discourse codes, grounds codes, conceptual quality codes, and the overarching approach. These decisions represent examples of the importance of connecting theoretical commitments about argumentation with discipline-specific features and analytic frameworks.

Analysis of Relationships Between Discourse Episodes and Constituent Comments

The analysis focuses on the relationships between the oppositional level of a discourse episode and the types of constituent moves, grounds quality of those constituent moves, and conceptual quality of those constituent moves. We employ Cochran–Mantel–Haenzel chi-square analyses based on table scores to determine the significance of the relationship between the *discourse move* of a comment and the *grounds quality* or *conceptual quality* of that comment. Cochran–Mantel–Haenzel chi-square tests control for the effect of the different discussion groups. The Row Score Mean Differ value is generally used for most of the analyses rather than the Nonzero Correlation value or the General Association value because the *discourse move* variable is a nominal variable, whereas the *grounds quality* and *conceptual quality* variables are ordinal (Stokes, Davis, & Koch, 2001).

Summary of Coding Method

To summarize, the framework first scores the individual comments in terms of *discourse move*, *grounds quality*, and *conceptual quality*. Each comment is coded in the context of its parent comment:

- *Discourse move* is a categoric (nominal) code representing the comment's role or intended role in a co-constructed dialogic argument, as shown in Table 1. Sample *discourse move* categories include: *claim*; *rebuttal against grounds*; *organization of participation*; and *off-task* comments.
- *Grounds quality* is an ordinal code with four levels representing the quality of grounds included with a comment, as shown in Figure 1. The ordinal levels of *grounds quality* include: no grounds (*level 0*); explanation only (*level 1*); evidence only (*level 2*); and evidence and an explanation or coordinating multiple pieces of evidence or as grounds (*level 3*).
- *Conceptual quality* is an ordinal code with four levels representing the conceptual quality of the scientific subject matter, as shown in Figure 2. The ordinal levels of *conceptual quality* include: non-normative (*level 0*); transitional (*level 1*); normative (*level 2*); and nuanced (*level 3*) use of content.

After coding the individual comments, the discussion is parsed into discourse episodes based on the second-order comments. This is a completely mechanical process based on responses to the initial seed claims. The framework then assigns a *level of opposition* code to each discourse episode (Table 4) based on the *discourse move* codes (Table 1) of the constituent comments. The *level of opposition* code is an ordinal code with six levels that measures the amount of conflict within the episode from a structural (e.g., presence of counterclaims, rebuttals against thesis, or rebuttal against grounds) rather than a conceptual perspective. These *levels of opposition* include: non-oppositional episodes (*level 0*); argumentation with claims or counterclaims but no grounds or rebuttals (*level 1*); argumentation with claims or counterclaims and grounds but no rebuttals (*level 2*); argumentation with grounds and a single rebuttal (*level 3*); argumentation with multiple rebuttals (*level 4*); and argumentation with multiple rebuttals and at least one *rebuttal against grounds* (*level 5*).

Context and Data for the Current Study

This study analyzes eight randomly chosen online personally seeded discussions involving a total of 84 students from four classes of eighth grade students. The *personally-seeded discussion* system is a customized asynchronous online discussion forum that is currently embedded within the Web-based Inquiry Science Environment (WISE) project *Thermodynamics: Probing Your Surrounding* (see http://wise.berkeley.edu). Each online discussion involves approximately five pairs of students. The students in this study worked on this project over a period of 2 weeks under the supervision of one experienced teacher who has worked extensively with the WISE research group in piloting other WISE projects. He had not worked with our argumentation approach, however, prior to this study. The public school is located in a diverse city and has an even distribution of boys and girls. The classes are typical eighth grade physical science classes, labeled neither "honors" nor "remedial." The teacher's curriculum for that semester focused on physical science with substantial emphases on thermodynamics and light. The curriculum involved substantial small group work where students conducted experiments, negotiated ideas, and drew conclusions based on evidence they had gathered. Prior to this study, however, the students had not explicitly studied dialogic argumentation within the curriculum.

The *Thermodynamics: Probing Your Surroundings* project consists of eight activities (Fig. 3). In activities 1–5 students collect real-time data about the temperatures of objects found inside the classroom and explore interactive simulations dealing with such ideas as heat transfer, thermal conductivity, and thermal sensation. As students work through these activities they are prompted to record the data they gather and describe the observations they make using the WISE note feature. Together, these five activities are designed to provide students with the empirical data and other scientific ideas needed to stimulate productive argumentation during the *personally-seeded discussions* described in our theoretical overview. We have provided more detailed information about the project, the personally-seeded discussions, and the theoretical rationale for our approach in prior publications (Clark & Sampson, 2005, 2006a, in press).

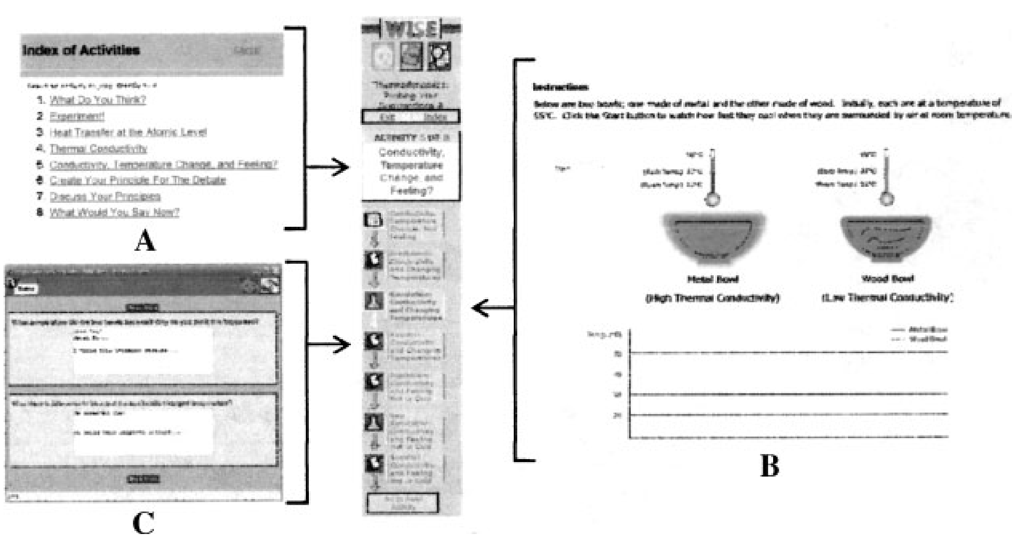

Figure 3. In the *Thermodynamics: Probing Your Surrounding* project, students (A) work through a series of activities, (B) explore interactive simulations, and (C) record their observations and ideas in the form of online notes before participating in the *personally-seeded discussions*.

In activity 6, students are asked to develop a principle that explains why objects that have been sitting in the same room for long periods of time often feel different. To scaffold students in this task and to ensure that they articulate their ideas clearly and focus on the salient issues of the problem, students use the *PrincipleMaker* interface. This interface allows students to use a pull-down menu format to create a principle from sentence fragments (Fig. 4). The predefined phrases and elements include components of inaccurate principles that students typically use to describe heat, thermal equilibrium, and thermal conductivity that were identified through the misconceptions and conceptual change literature (Albert, 1978; Clough & Driver, 1985; Erickson, 1979; Erickson & Tiberghien, 1985; Harrison, Grayson, & Treagust, 1999; Hewson & Hamlyn, 1984; Jones, Carter, & Rua, 2000; Lewis & Linn, 1994; Rogan, 1985; Shayer & Wylam, 1981; Slone, Tredoux, & Bokhorst, 1996) and an earlier thermodynamics curriculum development project (Clark, 2000, 2001; Lewis, 1996; Linn & Hsi, 2000). This process serves multiple purposes. First, the pull-down format ensures that the students' conceptions of a phenomenon focus on the salient issues and are well elaborated so that other students notice differences and want to discuss them. Second, the pull-down menu format enables the discussion software to differentiate between students' principles so that students can be automatically assigned to a discussion forum with other students who have constructed different principles to explain the same phenomenon.

Once the students have submitted their principles, they move on to activity 7. During this activity students participate in an asynchronous online discussion where they are encouraged to propose, support, critique, evaluate, and revise ideas. To foster argumentation in this type of context, we have designed the personally-seeded discussion software to set up discussion forums that consist of three to five pairs of students who have created different principles to explain the same phenomenon. This ensures that students are exposed to alternative interpretations of a given phenomenon, which, according to Osborne et al. (2004), is a ‘‘pedagogical strategy that will both initiate and support argumentation’’ (p. 997). Once students enter a discussion forum, they see the principle they have created along with the principles created by two to four other pairs of students (Fig. 5). The students are then directed to read and critique each principle with the goal of developing a shared understanding.

This type of activity moves beyond simply telling students to evaluate one another's ideas. By asking the students to construct preliminary principles before joining the discussion we provide students with an opportunity to develop ownership over their ideas. This ownership provides

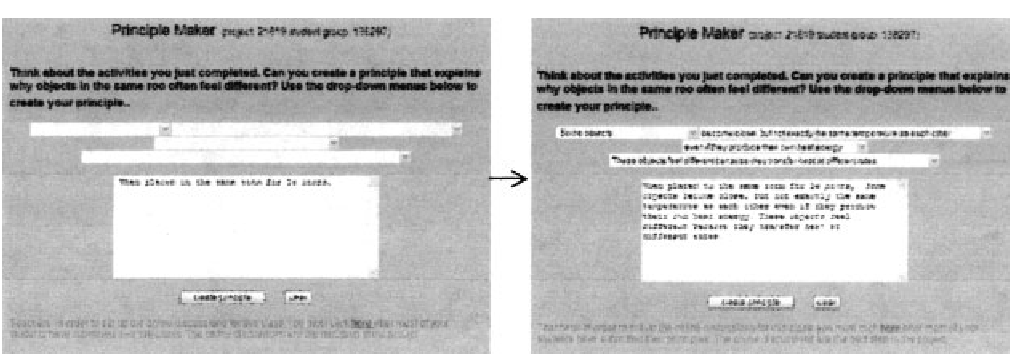

Figure 4. The *PrincipleMaker* interface. Students use a pull-down menu to make a principle from sentence fragments that include common misconceptions about heat transfer, thermal equilibrium, thermal conductivity, and thermal sensation.

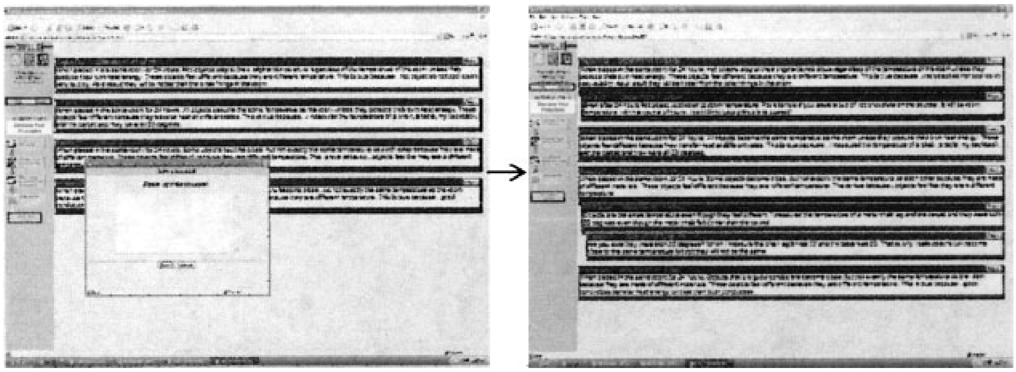

Figure 5. Personally-seeded discussions use the principles created by the students by means of the *PrincipleMaker* interface as seed comments. Students are automatically organized into discussion groups with students who have created different principles. When a student replies to a comment, their response is placed under the parent comment and indented.

students with intrinsic motivation to defend their principle. In addition, by asking students to agree upon a single answer they must engage with other students' ideas so that they can either weed out inaccurate ideas or combine ideas from their differing principles. Thus, to come to a consensus, students need to compare their principles, pay attention to one another as they defend and question each other, and revise their final answer accordingly. In short, they need to engage in argumentation to be successful. Thus, this activity structure goes beyond typical small group work. Rather then viewing small group work as an opportunity to divide up the labor to finish a task quicker, or as an opportunity to rely on a more knowledgeable peer (Cohen, 1994; Linn & Burbules, 1993), students must engage in genuine collaboration and consensus-building. Finally, in activity 8 students are asked to reflect on their experiences and what they have learned by participating in the project.

Results and Discussion

What can the proposed analytical framework tell us about the argumentation occurring within this environment and about students' dialogic argumentation in general? Through this process, what can we determine about the reliability, affordances, and constraints of the proposed analytic framework?

Relationships Between Discourse Move, Grounds, and Conceptual Quality

The discussions analyzed for this study included 334 student comments. When we applied the coding scheme to these comments, 264 were eligible to receive a grounds code (Fig. 6). Only 51% of these comments included some type of grounds (*grounds quality level 1* or higher). When students did include grounds, they relied on an explanation without evidence (*grounds quality level 1*) to support their ideas 47% of the time. Moreover, when students included evidence to justify their ideas, they rarely attempted to coordinate multiple pieces of evidence (*grounds quality level 3*); instead, they relied on rather simple justifications. Given the fact that the personally-seeded discussions are designed to encourage students to support and challenge ideas using evidence, the students' lack of attentiveness to this important aspect of scientific argumentation is problematic.

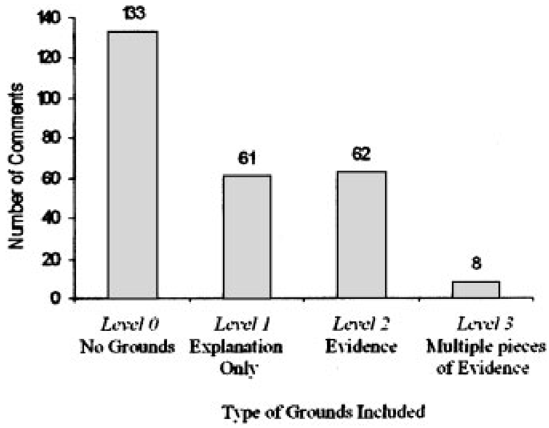

Figure 6. Frequency of comments in terms of their use of grounds.

However, when students' use of grounds is examined in terms of specific *discourse moves*, a different picture emerges. This analysis suggests that students use grounds differently depending on the *discourse move* of their comment. Figure 7 shows the results of this analysis (Fig. 7 shows *grounds quality levels 2 and 3* collapsed into a single category to facilitate analysis due to low frequency counts for *grounds quality level 3*). When students challenge the ideas of others, they frequently include grounds (*grounds quality level 1* or higher as opposed to 0) to support their comments. Seventy-three percent of the comments coded as a *rebuttal against thesis* include grounds and 74% of the comments coded as *rebuttal against grounds* include grounds. In contrast, when students support another student-pair, they only include grounds to justify their comments 31% of the time. Students usually do not provide grounds when attempting to strengthen their position in terms of accuracy or validity in response to a rebuttal; only 47% of the comments coded as a *clarification in response to a rebuttal* include grounds. These data indicate that students tend to support their comments by including higher quality grounds when they challenge the ideas presented by another student-pair than they do when offering support for an idea.

To determine whether this relationship between the *discourse move* of a comment and the *grounds quality* of that comment is significant, a Cochran–Mantel–Haenzel chi-square analysis

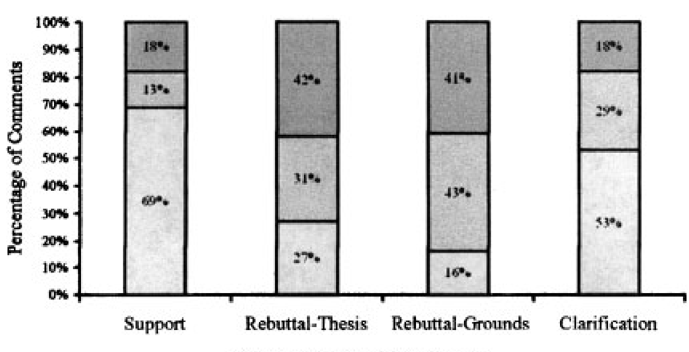

Figure 7. Percentage of comments at each level of *grounds quality* based on *discourse move.* White bars: no grounds (level 0); lightly shaded bars: explanation only (level 1); darkly shaded bars: evidence (levels 2 and 3).

based on table scores was conducted (as discussed in the section "Analysis of Relationships Between Discourse Episodes and Constituent Comments"). The possible *discourse move* codes included in the analysis were *rebuttal against thesis*, *rebuttal against grounds*, *support*, and *clarification in response to a rebuttal.* The test is significant, with a Row Score Mean Differ (3) value of 28.69 and $p < 0.01$. Based on this analysis, we are able to reject the null hypothesis that there is no relationship between these two variables, controlling for the effect of discussion group. Taken together, these results support the assertion that there is a significant relationship between the *discourse move* of a comment and *grounds quality*; students are more likely to provide grounds to support their statements when they are challenging the ideas of other student-pairs.

This analysis also indicates that students use grounds differently depending on whether or not their ideas are challenged by another student-pair. Students only include grounds 13% of the time when they are responding to a supporting comment. On the other hand, when students respond to a *rebuttal against thesis*, they include grounds 67% of the time and they include grounds 58% of the time when responding to a *rebuttal against grounds* (Fig. 8). To evaluate the significance of this relationship, a Cochran–Mantel–Haenzel chi-square analysis based on table scores was conducted to evaluate the hypothesis that there is a relationship between the *discourse move* of a comment and the *grounds quality* of the comments that reply to it. *Grounds quality levels 2 and 3* were again collapsed. The test is significant, with a Row Mean Score Differ (3) value of 7.92 and $p = 0.05$. Thus, these results support the assertion that students are more likely to include higher-quality grounds responding to a rebuttal than when responding to a supporting or a clarifying comment.

Relationships Between Levels of Opposition Within Episodes, Grounds, and Conceptual Quality

By splitting the discussions into discourse episodes based on the second-level comments (as discussed in the section "Parsing the Discussions into Discourse Episodes"), we identified a total of 126 discourse episodes. (Note that this system of parsing considers initial seed comments as

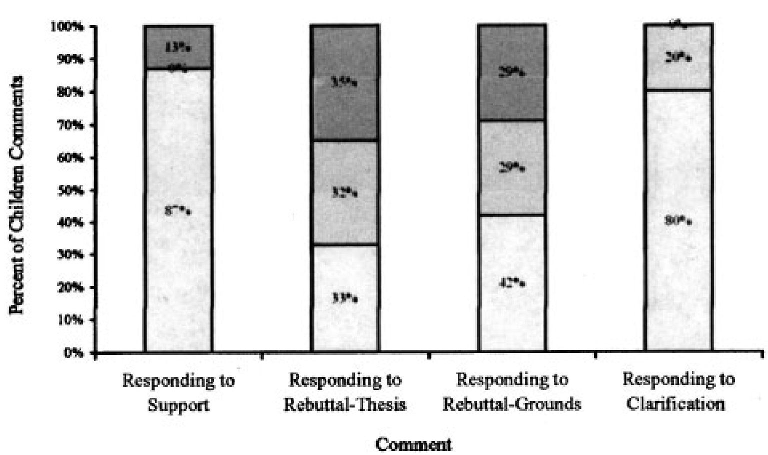

Figure 8. Comments at each level of *grounds quality* in response to comments of *various discourse moves.* Darkly shaded bars: children's comments with evidence as grounds (levels 2 and 3; lightly shaded bars: children's comments with an explanation as grounds (level 1); white bars: children's comments with no grounds (level 0).

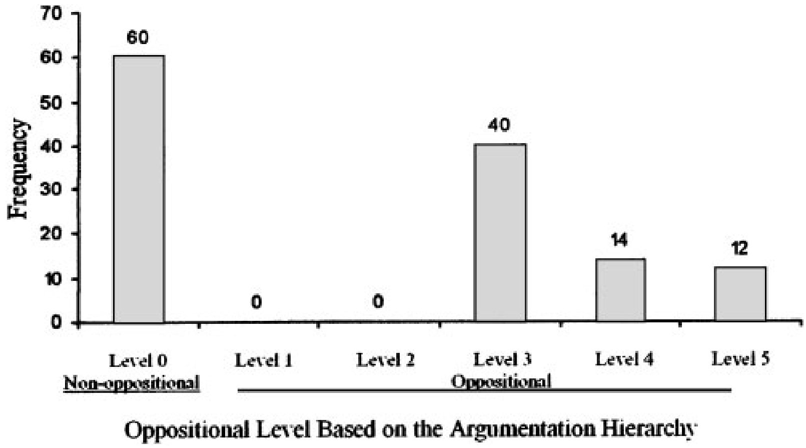

Figure 9. Frequency of discussion episodes at each *level of opposition.*

members of multiple episodes.) Of these episodes, 66 qualified as oppositional episodes and 60 did not (Fig. 9). Most non-oppositional episodes (*oppositional level 0*) tended to be relatively short (2.1 comments on average) and unsophisticated. Students apparently tended to accept what was written in the claim and move on to respond to other comments. The oppositional episodes tended to be longer. *Oppositional level 3* episodes involved 2.8 comments on average, *oppositional level 4* episodes involved 5.1 comments on average, and *oppositional level 5* episodes involved 5.9 comments on average (if two outlier episodes with 27 and 24 comments are excluded). The next phase of the analysis examines the relationship between the *level of opposition* found within the episodes and the *grounds quality* and *conceptual quality* of the comments found within these episodes.

Overall, the discourse episodes that included higher levels of opposition also had higher levels of *grounds quality.* Only 24% of the comments in non-oppositional episodes (*oppositional level 0*) included some type of grounds. In terms of oppositional episodes, 49% of comments in episodes with a single rebuttal against thesis (*structural quality level 3*) included grounds, 61% of comments in episodes with multiple rebuttals (*oppositional level 4*) included grounds, and 67% of comments in episodes with multiple rebuttals and at least one rebuttal against grounds (*oppositional level 5*) included grounds (Fig. 10). These observations also support the assertion that students tend to generate and incorporate stronger grounds to support their ideas when they participate in argumentation characterized by a high frequency of challenges.

To evaluate the hypothesis that there is a relationship between the *level of opposition* found with an episode and the *grounds quality* of the constituent comments, a Cochran–Mantel–Haenzel chi-square analysis based on table scores was conducted. *Oppositional levels 1 and 2* were omitted due to low frequency. *Grounds quality levels 2 and 3* were collapsed as with the early analyses. The test was conducted to control for the effect of the different discussion groups. The test was significant, with a Nonzero Correlation (1) value of 24.38 and $p < 0.01$. The Nonzero Correlation value was used in this analysis rather than the Row Mean Score Differ value because both variables are ordinal (Stokes et al., 2001). Students tend to provide higher-quality grounds to support their ideas when they are participating in argumentation characterized by higher levels of opposition.

The *oppositional level* of a discourse episode also appears to be related to the average *conceptual quality* of the comments within the episode (Fig. 11). In non-oppositional episodes

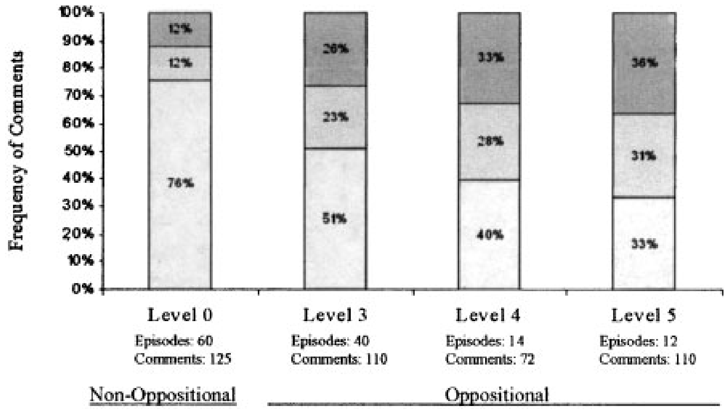

Figure 10. Percentage of the total comments within each *level of opposition* in terms of grounds. *Grounds quality* included as part of comment. Darkly shaded bars: evidence (levels 2 and 3); lightly shaded bars: explanation only (level 1); white bars: no grounds (level 0).

(*oppositional level 0*), only 21% of the comments were scored as either conceptually normative or nuanced (*conceptual quality levels 2 and 3*). In oppositional episodes with a single rebuttal (*oppositional level 3*) the percentage rose to 41%; in episodes with multiple rebuttals (*oppositional level 4*) it climbed to 45%; and, in episodes with multiple rebuttals and at least one *rebuttal against grounds* (*oppositional level 5*), 60% of the comments were either normative or nuanced (*conceptual quality levels 2 and 3*).

To evaluate the significance of this relationship, a Cochran–Mantel–Haenzel chi-square analysis based on table scores was used to evaluate the hypothesis that there is a relationship between *the level of opposition* found within episodes and the *conceptual quality* of the comments that occur within these episodes. The analysis involved four *oppositional levels*, including non-oppositional episodes (*level 0*) and oppositional episodes *levels 3, 4, and 5* and four levels of

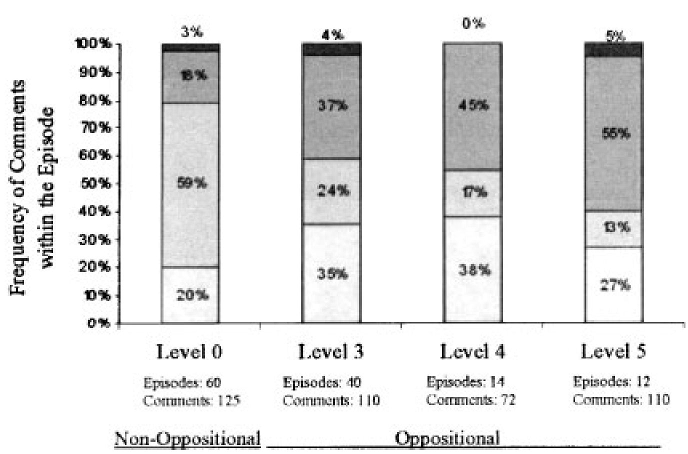

Figure 11. Percentage of comments within each *oppositional level* in terms of *conceptual quality. Conceptual quality* of the comment: Darkest bars: nuanced; darker bars: normative; lighter bars: transitional; white bars: non-normative.

conceptual quality, including non-normative (*conceptual quality level 0*), transitional (*conceptual quality level 1*), normative (*conceptual quality level 2*), and nuanced (*conceptual quality level 3*). Because both of these variables are ordinal, the Nonzero Correlation value was used in this analysis (Stokes, Davis & Koch, 2001). The test is significant when run without controlling for discussion group with a Nonzero Correlation (1) value of 3.93 and $p = 0.05$. When controlling for discussion group, however, the test loses significance, with a Nonzero Correlation (1) value of 1.59 and $p = 0.21$. These two tests suggest that students' comments tend involve a higher conceptual quality of subject matter when they are participating in argumentation of higher structural quality, but that further investigation is necessary to rule out inter-discussion group influences on this finding.

What Does the Proposed Analytic Framework Suggest About Students' Argumentation Within this Environment?

The proposed framework, which integrates issues of *discourse move* and *levels of opposition* with issues of *grounds quality* and *conceptual quality*, provides new insights into this environment. Earlier analyses of the environment (Clark & Sampson, in press) looked specifically at the nature of the argumentation using Toulmin's model of argument structure in order to facilitate comparisons to the larger body of work in science education devoted to promoting and supporting argumentation (e.g., Bell & Linn, 2000; Erduran et al., 2004; Jimenez-Aleixandre et al., 2000; Kelly et al., 1998; Osborne et al., 2004). That study suggested that personally-seeded discussions are an effective way to encourage students to justify their ideas and challenge the ideas of others as indicated by students' use of grounds and rebuttals. The application of the proposed analytic framework suggests that both the principle-creation interface, which makes the discussion group's thinking visible and articulate, as well as the automated discussion-group creation, which places students into discussion groups with student-pairs who created different principles, facilitates high levels of oppositional discourse. Not only is this type of opposition desirable in and of itself in terms of providing opportunities for students to learn how to justify, evaluate, and revise scientific ideas, but it is also related to how students use grounds and the conceptual quality of their ideas.

The effectiveness of this online environment as a way to foster this type of discourse is encouraging, especially in light research suggesting that students rarely have an opportunity to discuss, evaluate, and debate the processes and products of inquiry (Newton, Driver, & Osborne, 1999; Simon, Osborne, & Erduran, 2003). As Driver et al. (2000) suggested, ''the major barrier to developing young people's skills of argumentation in science is the lack of an opportunity offered for such activities within current pedagogical practices'' (p. 308). Therefore, the success of this environment is particularly attractive given the relatively low levels of teacher training necessary for online environments such as the WISE environment (Linn, Davis, & Bell, 2004), in comparison to the training necessary for the teacher to promote and support productive argumentation directly (Osborne et al., 2004). Note that the teacher in this study had run many WISE projects prior to this study, but that he had not worked with our argumentation approach before. Obviously, online environments are not substitutes for well-trained teachers, but online environments do offer potentially valuable supplemental affordances. Online environments may in fact even help import and model processes of argumentation for the teachers as well as the students. Certainly there is room for improving the personally seeded discussion approach, but the progress and future potential seem promising.

What Can We Say About the Reliability and Appropriateness of the Proposed Analytic Framework?

In assessing the reliability and appropriateness of the proposed analytic framework, we first discuss the interrater reliability of the coding schemes for *discourse move*, *conceptual quality*, and *grounds quality* for individual comments. To establish the interrater reliability of the coding scheme, all eight discussions were independently coded by the two authors and compared. We then examine the appropriateness of the episode parsing scheme and the way we code and score the oppositional quality of argumentative discourse.

Interrater reliability. In terms of *discourse moves*, interrater reliability was initially 93%. The largest category of difference between the two coders involved distinguishing *off-task* comments versus *support* comments. We increased interrater reliability to 94% (Cohen's kappa = 0.91) by refining the definition of *support* to include comments that express appreciation for prior support (rather than labeling such comments as *off-task*). The remaining 6% of differences were resolved through discussion. This remaining variance is minimal considering that a Cohen's kappa value of 0.70 is typically used as a benchmark for adequate interrater reliability (Bakeman & Gottman, 1986; Fleiss, 1981).

The interrater reliability for the coding of *conceptual quality* is 94% (Cohen's kappa = 0.93) and the remaining 6% of variance between the coders was resolved through discussion. The largest category of disagreement between the coders involved several instances in which the *conceptual quality* of an individual facet within a comment was scored differently. As discussed in the section "Coding the Conceptual Quality of a Comment," each coder must first determine how many non-normative, transitional, and normative facets are included as part of the entire comment using conceptual facet tables (see Table 3). After coding the individual facets of a claim, the overall *conceptual quality* of a comment is then determined based on the frequency of non-normative, transitional, and normative facets found within the entire comment using the flowchart in Figure 2. Therefore, the overall *conceptual quality* score assigned to a comment will occasionally vary between coders because of a difference in the coding of an individual facet. To address this issue, we revised the conceptual facet tables to reduce ambiguity and improve the future overall reliability of the coding scheme in terms of *conceptual quality*.

Finally, the initial interrater reliability for the coding of *grounds quality* was 86% (Cohen's kappa = 0.79). The largest category of disagreement involved several instances in which one author assigned the *grounds quality level 1* code (explanation only), whereas the other author assigned the *grounds quality level 2* code (evidence). By refining the evidence definition, the interrater reliability climbed to 92% (Cohen's kappa = 0.91). The remaining differences were resolved through discussion. Similar to other studies, we believe that this remaining 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 (e.g., Erduran et al., 2004; Kelly et al., 1998).

Appropriateness of the episode parsing scheme. The episode parsing scheme is arbitrary and mechanical, resulting in 100% interrater reliability. The trade-off in choosing to arbitrarily define an episode by the second-order comments is that this definition might not truly capture single-discourse episodes. Instead, several independent discourse episodes might occur within an episode defined by the second-order comments. The practice of defining episodes based on the second-level comments seems appropriate, however, given the small degree of branching that occurs within the episodes. We can quantify this branching by comparing the total number of comments in an episode with the length of the longest continuous chain of comments within that episode. If there is no difference, then the episode is simply one long chain of comments. On the other hand, if a discussion episode consists of several comments replying to the same parent comment, the difference between the longest continuous chain and the total number of comments will be large, potentially indicating multiple discussions taking place within an episode.

The difference between mean total number and mean longest continuous chain is about 0.1 comment for non-oppositional and oppositional episodes of *structural quality level 3* or less. The difference for *structural quality level 4* arguments is 0.2 comment. The majority (10 of 12) *structural quality level 5* episodes approach a gray area, where the mean number of comments is 5.9 and the mean longest continuous chain is 4.2 episodes. The other two *structural quality level 5* episodes are more clearly problematic in terms of the branching issue. One of these episodes involves 24 total comments and the longest continuous chain involves 9 comments. The other episode involves 27 total comments and the longest continuous chain involves 7 comments. In the future it might be more appropriate to add an additional clause to the parsing scheme that specifies subdivision of episodes of this size into multiple episodes representing each of the substantial branches. Further analysis of extended episodes will be required to resolve this particular issue. Overall, however, this issue applied to only 2 episodes out of 126 and the parsing method for defining discourse episodes proved appropriate for the vast majority of the episodes.

Conclusions

Based on this study, the proposed analytic framework appears to provide a useful set of tools for analyzing the relationships between levels of opposition, discourse moves, use of grounds, and conceptual quality. Furthermore, the results suggest that the framework is reliable for trained coders despite its complexity. The framework builds on the work of many theorists and viably evaluates students' dialogic argumentation practices and conceptual understanding of thermodynamics. Many other coding schemes have focused exclusively on argumentation structure, particularly those schemes drawing on Toulmin's work. This framework integrates issues of *discourse move* and *levels of opposition* with issues of *grounds quality* and *conceptual quality.*

The framework still has room for improvement and growth, however. Although the framework can relate the *conceptual quality* of individual comments to the overall *level of opposition* of the episode in which it occurs, the framework would be much more powerful if it could trace the evolution and trajectory of conceptions within an episode or across a full discussion. Several good frameworks exist with a focus on patterns and trajectories of interaction in dialogic argumentation (e.g., Baker, 2003; Hogan et al., 2000; Leitão, 2000; Weinberger & Fischer, 2006). If we could integrate characterizations of the patterns or trajectories of interaction into the framework, we could characterize the discussion as a whole, as opposed to breaking it into its constituent episodes, and we could track individual student-pairs and the evolution of their contributions over time.

A second area for future work involves adapting the framework for application in offline settings. Issues of parsing episodes and establishing referents between comments are more complicated, but the rest of the framework should transfer directly considering that the framework builds on the work of researchers from both online and offline contexts. Establishing discourse episodes offline would involve determining when a new major topic begins. Assigning comments to the comments to which they refer would also be important, rather than simply considering the comments chronologically, because sometimes students refer to earlier comments when new ideas arise. A rubric for making these decisions would be necessary but certainly manageable. The key would involve importing the methodologies for parsing and assigning referents outlined by Erduran et al. (2004), because their methodologies are designed for standard classroom contexts.

Although this study did not focus on offline contexts, the framework should theoretically adapt well to standard classroom contexts.

A third area, related to the second, involves exploring the adaptability of the framework to other epistemological stances regarding argumentation. As discussed earlier, our coding of structural argumentation features draws heavily on the work outlined by Erduran et al. (2004), but with significant adaptations to fit the epistemological nature of argumentation in the context that we study. There is, however, flexibility in the system to adapt to other epistemological stances, such as that outlined by Erduran et al. (2004) with regard to the nature and definition of rebuttals. Those epistemological stances could be exchanged into our framework if researchers believe that those stances more closely aligned with their research contexts. Future work in this area might examine the types of flexibility available for adapting the framework to other epistemological stances regarding argumentation as dictated by context or pedagogical goals.

A fourth area for future work involves expanding the coding schemes to more specifically characterize the sources students use to support their ideas or to measure the causal versus descriptive nature of students' explanations. Although the current version of the analytical framework represents a solid move toward integrating the conceptual content of student argumentation, further work will obviously improve upon these schemes. It is our hope that this framework will continue to evolve and provide a tool for other researchers interested in the interplay of *discourse move*, *grounds quality*, and content *conceptual quality* in co-constructed dialogic argumentation.

In conclusion, the results of this study demonstrate that levels of opposition, grounds quality, and conceptual quality show significant interrelationships. By focusing on all three of these issues, rather than solely on the structure of the arguments generated by students, we could substantially enhance the argumentation that takes place within science classrooms. The proposed analytical framework provides a potential tool for further analyzing and bootstrapping these interrelationships. With further improvements by our research group and other researchers, we hope that the proposed analytical framework will augment research and development of online and offline science learning environments to better scaffold students' understanding of scientific argumentation and its role in science and inquiry.

References

Abell, S.K., Anderson, G., & Chezem, J. (2000). Science as argument and explanation: Exploring concepts of sound in third grade. In J. Minstrell & E.H. Van Zee (Eds.), Inquiry into inquiry learning and teaching in science (pp. 100–119). Washington, DC: American Association for the Advancement of Science.

Albert, E. (1978). Development of the concept of heat in children. Science Education, 62, 389–399.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. Washington, DC: Author.

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational contexts of confronting cognitions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 1–25). Dordrecht: Kluwer Academic.

Andriessen, J., Erkens, G., Van de Laak, C., Peters, N., & Coirier, P. (2003). Argumentation as negotiation in electronic collaborative writing. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 79–115). Dordrecht: Kluwer Academic.

Bakeman, R., & Gottman, J. (1986). Observing interaction: An introduction to sequential analysis. Cambridge, UK: Cambridge, University Press.

Baker, M. (2003). Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 47–78). Dordrecht: Kluwer Academic.

Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (in press). Rainbow: A framework for analyzing computer-mediated pedagogical debates. International Journal of Computer Supported Collaborative Learning.

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, 797–818.

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

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. International Journal of Science Education, 11, 514–529.

Chinn, C.A., & Brewer, W.F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. Journal of Research in Science Teaching, 35, 623–654.

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. (in press). Longitudinal conceptual change in students' understanding of thermal equilibrium: Conceptual restructuring from fragmentation toward coherence. Cognition & Instruction.

Clark, D.B., & Linn, M.C. (2003). Scaffolding knowledge integration through curricular depth. Journal of Learning Sciences, 12, 451–494.

Clark, D.B., & Sampson, V. (2005). Analyzing the quality of argumentation supported by personally-seeded discussions. Paper presented at the Computer Supported Collaborative Learning Conference, Taipei, Taiwan.

Clark, D.B., & Sampson, V. (2006a). Characteristics of students' argumentation practices when supported by personally-seeded discussions. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Clark, D.B., & Sampson, V. (2006b). The quality of argumentation supported by personally-seeded discussions. In T. Koschmann, T-.W. Chan, & D. Suthers (Eds.), Computer supported collaborative learning 2005. Mahwah, NJ: Lawrence Erlbaum.

Clark, D.B., & Sampson, V. (in press). Personally-seeded discussions to scaffold online argumentation. International Journal of Science Education.

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

Clough, E.E., & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: Bringing together scientific and personal views. The Physical Educator, 20, 176–182.

Cohen, E.G. (1994). Restructuring the classroom: Conditions for productive small groups. Review of Educational Research, 64, 1–35.

deVries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. Journal of the Learning Sciences, 11, 63–103.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84, 287–313.

Duschl, R. (in press) Quality argumentation and epistemic criteria. In S. Erduran & M. Jimenez-Aleixandre (Eds.), Argumentation in science education: Recent developments and future directions. New York: Springer.

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

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.

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

Erickson, G., & Tiberghien, A. (1985). Heat and temperature. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), Children's ideas in science (pp. 52–83). Philadelphia: Open University Press.

Erickson, G.L. (1979). Children's conceptions of heat and temperature. Science Education, 63, 221–230.

Erkens, G., & Janssen, J. (2006). Automatic coding of communication in collaboration protocols. Paper presented at the Proceedings of the 7th International Conference of the Learning Sciences, Bloomington, IN.

Erkens, G., Kanselaar, G., Prangsma, M., & Jaspers, J. (2003). Computer support for collaborative and argumentative writing. In E.D. Corte, L. Verschaffel, N. Entwistle, & J.v. Merriënboer (Eds.), Powerful learning environments: Unravelling basic components and dimensions (pp. 157–176). Amsterdam: Pergamon Elsevier.

Fabos, B., & Young, M.D. (1999). Telecommunication in the classroom: Rhetoric versus reality. Review of Educational Research, 69, 217–259.

Fischer, F. (2001). Gemeinsame Wissenskonstruktion. Analyse und Förderung in computerunterstützten Kooperationsszenarien [Collaborative knowledge construction. Analysis and facilitation in computer-supported collaborative scenarios]. Ludwig-Maximilians-Universität München, München.

Fleiss, J.L. (1981). Statistical methods for rates and proportions. New York: Wiley.

Harrison, A.G., Grayson, D.J., & Treagust, D.F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. Journal of Research in Science Teaching, 36, 55–87.

Hesse, F. (2007). Being told to do something or just being aware of something? An alternative approach to scripting in CSCL. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), Scripting computer-supported communication of knowledge: Cognitive, computational and educational perspectives.

Hewson, M.G., & Hamlyn, J. (1984). The influence of intellectual environment on conceptions of heat. European Journal of Science Education, 6, 245–262.

Hoadley, C., & Linn, M.C. (2000). Teaching science through on-line peer discussions: SpeakEasy in the knowledge integration environment. International Journal of Science Education, 22, 839–857.

Hogan, K., Nastasi, B., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. Cognition and Instruction, 17, 379–432.

Janssen, J., Erkens, G., Jaspers, J., & Kanselaar, G. (2006, June/July). Visualizing participation to facilitate argumentation. Paper presented at the 7th International Conference of the Learning Sciences, Bloomington, IN.

Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 205–226). Dordrecht: Kluwer.

Jermann, P., Soller, A., & Muehlenbrock, M. (2001). From mirroring to guiding: a review of state of art technology for supporting collaborative learning. Paper presented at the European Computer Supported Collaborative Learning Conference, Maastricht, NL.

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

Joiner, R., & Jones, S. (2003). The effects of communication medium on argumentation and the development of critical thinking. International Journal of Educational Research, 39(8), 861–971.

Jones, M.G., Carter, G., & Rua, M.J. (2000). Exploring the development of conceptual ecologies: Communities of concepts related to convection and heat. Journal of Research in Science Teaching, 37, 139–159.

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

Kirschner, P.A., Buckingham Shum, S., & Carr, C.S. (2003). Visualizing argumentation: software tools for collaborative and educational sense-making. London: Springer.

Kolodner, J., Schwarz, B., Barkai, R., Levy-Meumand, E., Tcherni, A., & Turbovsk, A. (1997). Roles of a case library as a collaborative tool for fostering argumentation. In R. Hall, N. Miyake, & N. Enyedy (Eds.), Proceedings of the 1997 computer-support for collaborative learning (CSCL '97). Hillsdale, NJ: Erlbaum.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. Science Education, 77, 319–337.

Kuhn, D., & Udell, W. (2003). The development of argument skills. Child Development, 74, 1245–1260.

Kuhn, L., & Reiser, B. (2005). Students constructing and defending evidence-based scientific explanations. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

Kuhn, L., & Reiser, B. (2006). Structuring activities to foster argumentative discourse. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Leitão, S. (2000). The potential of argument in knowledge building. Human Development, 43, 332–360.

Lewis, E.L. (1996). Conceptual change among middle school students studying elementary thermodynamics. Journal of Science Education and Technology, 5, 3–31.

Lewis, E.L., & Linn, M.C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. Journal of Research in Science Teaching, 31, 657–677.

Linn, M.C., & Burbules, N. (1993). Construction of knowledge and group learning. In K. Tobin (Ed.), The practice of constructivism in science education (pp. 91–119). Washington, DC: American Association for the Advancement of Science.

Linn, M.C., Davis, E.A., & Bell, P. (2004). Internet environments for science education. Mahwah, NJ: Lawrence Erlbaum.

Linn, M.C., & Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In P. Alexander & P.H. Winne (Eds.), Handbook of educational psychology. Mahwah, NJ: Lawrence Erlbaum.

Linn, M.C., & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: Lawrence Erlbaum.

Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. Instructional Science, 29, 127–153.

Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. In L. Mason (Eds.), Instructional practices for conceptual change in science domains. Learning and Instruction, 11, 305-329.

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21, 553–576.

Oestermeier, U., & Hesse, F. (2000). Verbal and visual causal arguments. Cognition, 75, 65–104.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in science classrooms. Journal of Research in Science Teaching, 41, 994–1020.

Osborne, J., Simon, S., & Erduran, S. (2002, September). Enhancing the quality of argumentation in school science. Paper presented at the Conference on Philosophical, Psychological, and Linguistic Foundations for Language and Science Literacy Research, University of Victoria, BC, Canada.

Pea, R.D. (1994). Seeing what we build together: Distributed multimedia learning environments for transformative communications. [Special Issue: Computer support for collaborative learning.] Journal of the Learning Sciences, 3, 285–299.

Perkins, D.N., Farady, M., & Bushy, B. (1991). Everyday reasoning and the roots of intelligence. In J.F. Voss, D.N. Perkins, & J.W. Segal (Eds.), Informal reasoning and education. Hillsdale, NJ: Erlbaum.

Rogan, J.H. (1985). The development of a conceptual framework of heat. Science Education, 10.

Roschelle, J., & Pea, R. (1999). Trajectories from today's WWW to a powerful educational infrastructure. Educational Researcher, 28, 22–25. 43.

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of the research. Journal of Research in Science Teaching, 41, 513–536.

Sandoval, W.A., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23, 23–55.

Sandoval, W.A., & Reiser, B.J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education, 88, 345–372.

Schauble, L., Glaser, R., Duschl, R., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. Journal of the Learning Sciences, 4, 131–166.

Schellens, T., & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. Computers & Education, 46, 349–370.

Schwarz, B., & Glassner, A. (2003). The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 227–260). The Netherlands: Kluwer Academic Publishers.

Schwarz, B.B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentative activity. Journal of the Learning Sciences, 12, 219–256.

Shayer, M., & Wylam, H. (1981). The development of the concept of heat and temperature in 10-13-year-olds. Journal of Research in Science Teaching, 18, 419–434.

Siegel, H. (1989). The rationality of science, critical thinking and science education. Synthese, 80, 9–42.

Simon, S., Erduran, S., & Osborne, J. (2002, April). Enhancing the quality of argumentation in school science. Paper presented at the annual meeting of the National Association for Research in Science Teaching, New Orleans, LA.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. International Journal of Science Education, 28, 235–260.

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: Routledge Falmer.

Slone, M., Tredoux, C., & Bokhorst, F. (1996). Decalage effects for heating and cooling: A cross-cultural study. Journal of Cross-Cultural Psychology, 27, 51–66.

Stokes, M.E., Davis, C.S., & Koch, G.G. (2001). Categorical data analysis using the SAS system (2nd ed.). New York: Wiley–SAS.

van Eemeren, F., Grootendorst, R., & Henkemans, A.F. (2002). Argumentation: Analysis, evaluation, presentation. Mahwah, NJ: Lawrence Erlbaum.

Veerman, A. (2003). Constructive discussions through electronic dialogue. In J. Andriessen, M. Baker, & D. Suthers (Eds.), Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments (pp. 117–143). Dordrecht: Kluwer Academic.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education, 46, 71–95.

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