

Chapter 10

Sequential Analysis of Scientific Argumentation in Asynchronous Online Discussion Environments*

Allan Jeong, Douglas B. Clark, Victor D. Sampson, and Muhsin Menekse

Abstract The national science standards call for increased focus on scientific argumentation in the classroom, and researchers have developed sophisticated online science learning environments to promote and support student engagement in scientific argumentation. Assessing the quality of scientific dialogic argumentation in these environments, however, has proven to be challenging. Existing analytic frameworks tend to assess scientific argumentation using the presence or absence of various types of comments (e.g., frequency of claims, rebuttals, and supporting statements) that do not fully convey the dynamic and dialogic nature of argumentation. In this chapter, we present a sequential analysis approach developed by Jeong (2005) that incorporates a coding scheme developed by Clark and Sampson (2007, 2008) to identify, visualize, and assess the dialogic processes of argumentation in online science learning environments in terms of transitional probabilities, transitional state diagrams, and other related measures. These measures include: (a) how and how often students respond to particular discourse moves (e.g., the probabilities that responses to claims are rebuttals *vs.* simple agreement *vs.* no response); and (b) how and to what extent observed response patterns produce extended chains of discourse moves that exhibit high levels of argumentation (claim → challenge → explain or amend claim). A sample analysis is presented to illustrate how this approach can also be used to assess how characteristics of the discourse environment affect the quality of argumentation and better understand the interplay between discourse environments and collaborative discourse.

*This work was partly funded by the National Science Foundation, grant 0334199, as part of the center for Technology Enhanced Learning in Science (TELS). Any opinions, findings, and conclusions or recommendations expressed in this study are those of the researchers and do not necessarily reflect the views of the National Science Foundation.

A. Jeong (✉)

Associate Professor, Florida State University, College of Education, Department of Educational Psychology & Learning Systems, Instructional Systems Program
e-mail: ajeong@fsu.edu

10.1 Introduction

Online learning environments designed to engage and support students in dialogic scientific argumentation provide excellent opportunities for students to propose, support, evaluate, critique, and refine ideas in a more productive manner. Over the last decade, a number of sophisticated environments have been developed to support students engaging in this type of knowledge-building or knowledge-validating discourse. Examples, among others, include *CONNECT* (e.g., deVries et al. 2002), *TC3* (e.g., Erkens et al. 2003), *DUNES* (e.g., Schwarz and Glassner in press), Virtual Collaborative Research Institute (e.g., Janssen et al. 2007), *ArgueGraph* (e.g., Jermann and Dillenbourg 2003), and the *personally-seeded discussions* within the Web-based Inquiry Science Environment (e.g., Clark 2004; Clark and Sampson 2007, 2008; Clark et al. 2009; Cuthbert et al. 2002). The multitude of approaches used to foster argumentation gives rise to complex and diverse assessment needs among researchers and an increasing interest in approaches for analyzing and assessing the nature or quality of dialogic scientific argumentation. To date, researchers have developed a broad range of methods that reflect various perspectives on argumentation, pedagogical goals, and curricular structures (see Clark et al. 2007 for a catalog of several of these methods). These methods tend to do an excellent job of providing overall ratings and observed frequencies of argumentative and collaborative interactions. However these frameworks tend not to provide information about the specific sequences of discourse moves produced in student exchanges – information that is needed to fully capture and computationally model the dynamic nature of argumentative discourse in CSCL (Jeong 2005).

For example, content analysis is one of the most common methods used in CSCL when analyzing learner interactions. In this method, researchers identify message categories and measure the frequency of messages observed in each category (Rourke et al. 2001). This approach generates results that are mainly descriptive rather than prescriptive in nature, reporting for example the frequencies of arguments, challenges, and explanations observed in a discussion. However, message frequencies provide little information that can be used to explain or predict how participants respond to given types of messages (e.g., argument → challenge *versus* argument → simple agreement), how response patterns are influenced by latent variables (e.g., message function, content, communication style, response latency) or exogenous variables (e.g., gender, personality traits, discussion protocols, type of task), and how particular response patterns contribute to observed differences in group performance on a desired outcome. Therefore, new approaches are needed to examine to what extent messages elicit responses based on what is said in conjunction with when, how, who, and why messages are presented, and whether or not the elicited responses help produce sequences of speech acts that support critical discourse (e.g., claim → challenge → explain) and group performance in decision making, problem-solving, and learning.

In this chapter, we integrate two complimentary methods that researchers can use in tandem to analyze and assess the nature of the interactions that take place

between students in CSCL environments that use asynchronous threaded discussion forums to engage students in scientific argumentation. The first method, developed by Clark and Sampson (2007, 2008), codes the nature of the discourse moves, the quality of the grounds used to support and challenge ideas, and the level of opposition that takes place between students as they propose, support, critique and refine ideas. The second method, sequential analysis (Bakeman and Gottman 1997) and the tools used to perform this types of analysis (Jeong 2005), captures the dynamic and dialogic nature of argumentation by measuring how and how likely students respond to various discourse moves of interest (e.g., the probabilities that responses to claims are rebuttals *vs.* simple agreement *vs.* no response), and how and to what extent these observed response patterns produce extended chains of discourse moves that reveal processes essential to producing high quality argumentation (claim→challenge→explain or amend claim). In the sections that follow, we will: (a) outline the sequence of steps, tools, and metrics used in each approach; (b) conduct a sample analysis that illustrates how these two methods can be used in tandem to compare and contrast various aspects of scientific argumentation; and (c) discusses implications and recommendations for researchers interested in using these approaches in tandem.

10.2 Steps, Tools, and Metrics Used in Each Approach

This section describes a procedure for coding the nature of the contributions made by the participants in an asynchronous discussion forum and the oppositional level of various discourse episodes (method 1) and the steps used to perform a sequential analysis of the argumentative discourse (method 2). This first method (Clark and Sampson 2007, 2008) consists of four major steps: (a) coding the discourse moves observed in individual postings/comments; (b) coding the grounds of a comment; (c) parsing the discussions into discourse episodes; and (d) scoring the level of opposition found within discourse episodes. This type of analysis enables a researcher to focus on specific episodes found within a discussion and provides a way to document the extent to which students question or challenge each other's ideas, how often they use grounds to support or challenge an idea, and the conceptual quality of students' ideas. Once this analysis is complete, a researcher can use the second method, sequential analysis, to identify response patterns measured in terms of the probabilities in which certain types of responses are elicited by given types of comments. This particular method, developed and refined by Jeong (2004, 2005), enables researchers to look at the discussion forum as a whole (across multiple episodes) to compare and identify similarities and differences in patterns of discourse produced between different groups under different conditions – patterns that might help to explain the observed number of times students question and challenge one another's ideas, use grounds to support or challenge ideas, and the quality of students' ideas.

10.2.1 Core Coding: Examining the Nature of Comments Found within Discourse Episodes

10.2.1.1 Coding the Discourse Moves of Individual Postings

The framework assigns a discourse move code to each comment based on the comment’s role in the discussion. In order 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, counter-claims, rebuttals), meta-organizational comments that help organize the interaction (which are typically overlooked in a structural analysis), and the occasional off-task interaction. The full list of *discourse move* comment codes is outlined in Table 10.1.

Table 10.1 Coding scheme for the discourse move of individual comments

Discourse move	Definition
Claim	The seed-comment principle or an assertion made.
Counter-claim	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 that indicates that: (1) they have changed their original claim; or (2) changed their viewpoint; or (3) have made a concession in response to comments (claims or rebuttals) made by another 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 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.

(continued)

Table 10.1 (continued)

Discourse move	Definition
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?”); (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!”).

10.2.1.2 Coding the Grounds of a Comment

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). We developed a series of binary decisions (see Fig. 10.1 flow chart) 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 for some comments (such as “organization of participation,” “query about meaning,” and “off-task”), these qualities simply do not apply. Coding of grounds is not the focus of the current chapter, but full detail about this aspect of the coding scheme is available in Clark and Sampson (2008).

10.2.1.3 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 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 developed through extensive prior conceptual change work measuring the longitudinal evolution of students’ conceptual ecologies (Clark 2000, 2006;

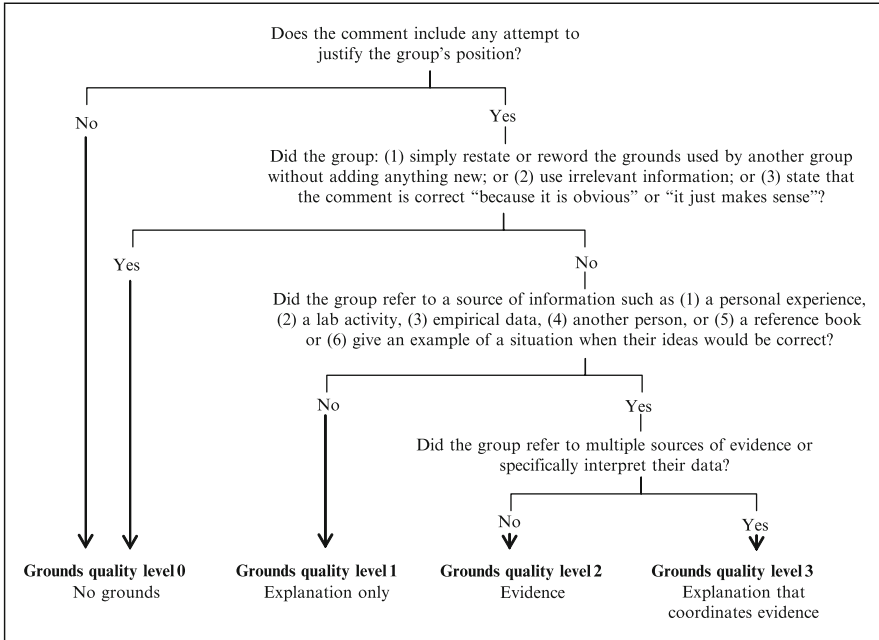


Fig. 10.1 Flow chart for coding grounds of a comment

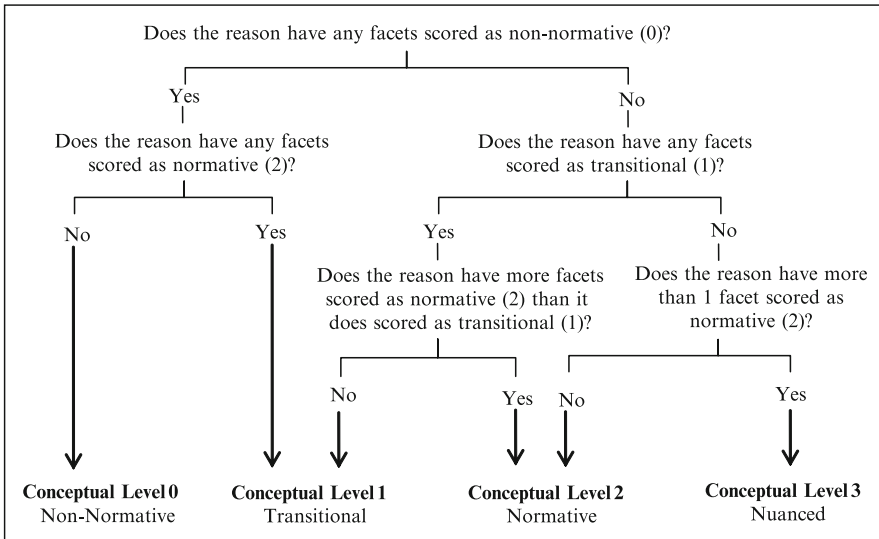


Fig. 10.2 Flow chart for coding the conceptual quality of a comment

Table 10.2 Example facets for coding conceptual quality of comment*Non-normative facets*

Metal objects are above/below ambient temperatures by a great difference

Size/thickness affects final temperature

Good conductors/insulators keep heat on the surface vs. keeping heat/cold inside

Transitional facets

All objects in the same room will reach close temp (but not the same) as surroundings

Insulators “block, trap, or allow a small amount of” heat or cold (act like barriers)

Normative facets

Objects in the same room become the same temperature

If more heat flows into an object than out of it, its temp rises (or reverse)

Clark and 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 in the flow chart (Fig. 10.2). The flow chart assigns an overall *conceptual quality* score based on the frequency of non-normative, transitional, and normative facets found within the entire comment (see Table 10.2 for examples). As with the discussion of grounds above, coding of conceptual quality is not the focus of the current chapter, but full detail about this aspect of the coding scheme is available in Clark and Sampson (2008).

10.2.1.4 Coding the Level of Opposition within Discourse Episodes

After coding the individual comments, the framework then codes the larger episodes of discourse within which the comments occur. The framework considers an episode to be defined by each second-level comment (including its parent claim and its children). The framework characterizes the amount of conflict or *level of opposition* that takes place within an episode using the hierarchy outlined in Table 10.3. 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 used to support this interpretation. On the other hand, low quality argumentation is either non-oppositional (*oppositional level 0*) or consists of only claims and counter claims which do not attempt to challenge the validity of the other participants interpretation of the phenomenon (*oppositional level 1*). This scheme adapts the hierarchy outlined in Erduran et al. (2004) by incorporating the expanded definition of rebuttals outlined in Clark and Sampson (2007, 2008).

10.2.2 Using Sequential Analysis to Identify Discourse Patterns in Argumentation

Sequential analysis (Bakeman and Gottman 1997) has been used to analyze and model sequential links between behavioral events to determine how likely one given event is followed by another given event. Jeong (2004, 2005) developed the

Table 10.3 The overall quality of the argumentation and level of opposition that takes place within an episode is 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 counter-claims with grounds but only a single rebuttal that challenges the thesis of a claim
Level 2	Argumentation involving claims or counter-claims with grounds but no rebuttals
Level 1	Argumentation involving a simple claim <i>versus</i> counter-claim with no grounds or rebuttals
Level 0	Non-oppositional

	+ARG	+BUT	+EXPL	+EVID	-ARG	-BUT	-EXPL	-EVID	Replies	No Replies	Givens	Reply Rate	% replies	% givens
+ARG	.02	.03	.25	.17	.00	.49	.00	.04	213	21	127	.83	.25	.10
+BUT	.00	.10	.05	.10	.00	.66	.06	.03	135	162	289	.44	.16	.23
+EXPL	.00	.02	.08	.15	.00	.67	.06	.02	52	64	112	.43	.06	.09
+EVID	.00	.00	.10	.13	.00	.71	.00	.06	31	50	84	.40	.04	.07
-ARG	.00	.48	.03	.02	.00	.02	.26	.19	174	21	124	.83	.20	.10
-BUT	.00	.61	.11	.02	.00	.08	.08	.09	157	185	328	.44	.18	.26
-EXPL	.00	.56	.13	.00	.00	.04	.17	.10	52	56	102	.45	.06	.08
-EVID	.00	.62	.05	.03	.00	.00	.15	.15	39	49	81	.40	.05	.06

Fig. 10.3 Transitional probability matrix produced by DAT

Discussion Analysis Tool (DAT) to compute the transitional probabilities between discourse moves observed in online debates. DAT has been used to produce transitional probability matrices to report, for example, the percentage of replies to stated arguments (ARG) that are challenges (BUT) *vs.* explanations (EXPL) *vs.* supporting evidence (EVID); and the percentage of replies to challenges that are counter-challenges *vs.* explanations *vs.* supporting evidence (see Fig. 10.3).

The matrix in Fig. 10.3 represents the message-response exchanges observed in an online debate. For example, the circled number indicates that 48% of all replies to the 124 opposing arguments (-ARG) were challenges (+BUT), for this group of students. The 124 opposing arguments (10% of all the discussion postings) elicited a total of 174 replies, approximately 20% of all the observed replies posted to the discussions. Only 21 of these 124 opposing arguments did not elicit any replies, and as a result, 83% of all opposing arguments elicited at least one or more replies.

DAT also produces a corresponding z-score matrix to identify and highlight transitional probabilities that are significantly higher/lower than expected

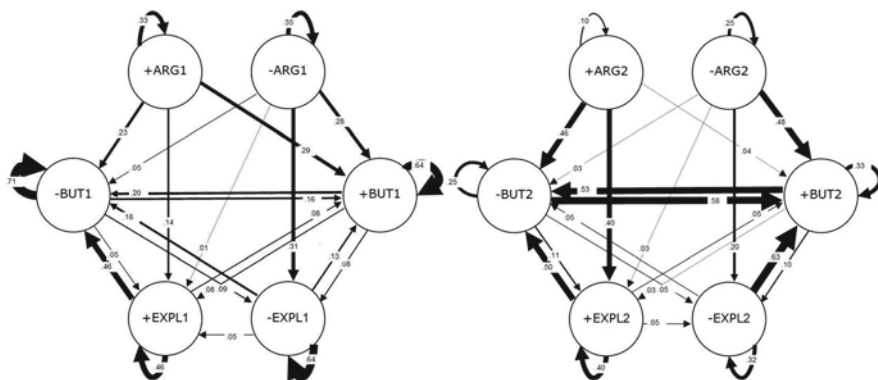


Fig. 10.4 Transitional state diagrams of response patterns produced by less intellectually open (*left diagram*) vs. more intellectually open students (*right diagram*)

probabilities. Probabilities that identify message-response sequences that can be considered to be behavioral “patterns” in an online debate. To visually and efficiently convey the complex data revealed in the transitional probability matrix, DAT converts the observed probabilities into transitional state diagrams (see Fig. 10.4). Potential differences in behavioral patterns between experimental groups—such as groups with students that have high vs. low in intellectual openness (Jeong 2007)—can be easily seen by juxtaposing state diagrams and observing the differences in the thickness of the links between events (signifying the strength of the transitional probabilities between given events). For example, a visual comparison of the two state diagrams in Fig. 10.4 shows that students that are more intellectually open (*right diagram*) exhibit a higher tendency to challenge one another’s arguments ($ARG \rightarrow BUT$) and counter-challenge one another’s challenges ($BUT \rightarrow BUT$) than students who are less intellectually open (*left diagram*).

To determine how an observed response pattern actually influences how often students post specific types of responses, DAT can be used to tabulate, for example, how many challenges are elicited by each argument, or how many explanations are elicited by each challenge. These scores can then be used to test for differences in the “mean response scores” – the mean number of challenges elicited per argument and the mean number of explanations elicited per challenge – between two or more experimental groups using statistical tests like the *t*-test and analysis of variance as demonstrated later in the case study.

10.3 A Sample Study and Analysis

To demonstrate the integration of the two methods described above, a case study was conducted to provide further insight into the findings of an earlier study (Clark et al. 2009). That study focused primarily on differences between two conditions in terms of differences in pre-post gains on the explanations that students constructed

before and after the discussions. A brief analysis was also conducted, however, on the discourse moves within the discussions of each condition using the base coding methods outlined earlier from Clark and Sampson (2008). The findings from the analysis of the discourse moves was suggestive in that study, but not conclusive. By integrating the sequential analysis component, we hope to provide further insight into the findings of Clark, D'Angelo, and Menekse.

10.3.1 Data Sample

This analysis focuses on five ninth-grade integrated science classes taught by the same teacher at a public high school in a large metropolitan area in the southwestern United States. This was the participant group from our first trial in the original study. The teacher was an experienced teacher, but he had not worked with the online environment employed in this study or our research group prior to this study. The classes were typical ninth grade integrated science classes, labeled neither "honors" nor "remedial." Prior to this study, the students had conducted various inquiry projects, but had not explicitly studied dialogic argumentation within the curriculum of the class. The students worked on the project for approximately six class periods. The public school is located in a diverse city and has a roughly even distribution of boys and girls. The district is 58% Non-Hispanic White, 29% Hispanic, 6% Black, 6% Asian/Pacific Islander, and 1.4% American Indian/Alaska Native. The district categorizes 27% of the student population as economically disadvantaged. In total, there were 147 students, 38 discussion groups, and 2,160 discussion comments.

10.3.2 Instructional Context

The *personally-seeded discussion* system that is the focus of this case study is a customized asynchronous online discussion forum embedded within a Web-based Inquiry Science Environment (WISE) project called *Thermodynamics: Probing Your Surrounding* (see <http://wise.berkeley.edu>). The *Thermodynamics: Probing your Surroundings* project consists of eight activities (see Fig. 10.5). In activities 1–5 students collected 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 worked through these activities they were prompted to record the data they gathered and describe the observations they made using the WISE note feature. We provide more detailed information about the project, the personally-seeded discussions, and the theoretical rationale for our approach in other publications (Clark 2004; Clark and Sampson 2007, 2008; Clark et al. 2009; Cuthbert et al. 2002).

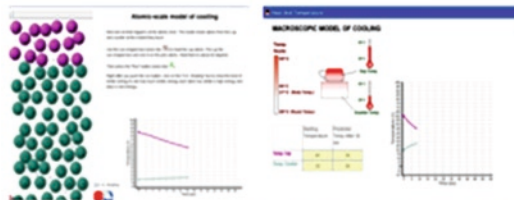
In activity 6, students were asked to develop a principle that explained why objects that have been sitting in the same room for long periods of time often feel different.

Activities in *Thermodynamics: Probing your Surroundings*

1 — What do you think? This activity introduces students to the driving question and elicits their ideas about thermodynamics.



2 — Experiment. This activity enables student to gather data about the temperature of different objects in the room using temperature probes.



3 — Heat Transfer at the Atomic Level. This activity enables students to explore simulations that illustrate the concept of heat transfer at the macroscopic and submicroscopic level.

4 — Thermal Conductivity. This activity introduces students to differences between thermal insulators and conductors by allowing them to explore an interactive simulation.



5 — Conductivity, Temperature Change, and Feeling. This activity enables students to explore differences between insulators and conductors in terms of how they feel and the rate in which the heat up or cool.

6 — Create Your Principle. This activity enables students to generate a principle that will explain their previous observations from a series of sentence fragments.

7 — Discuss your Principle Students enter a customized asynchronous discussion forum to discuss, support, challenge, and refine the principles created in activity 6.

8 — What do you think now? Students revisit the driving question and generate a final principle based on their experiences in the project.

Fig. 10.5 Overview of the activities in the *Thermodynamics: Probing Your Surrounding* project

To scaffold students in this task and to ensure that students 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 (see Fig. 10.6). 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 (e.g., Clough and Driver 1985; Erickson and Tiberghien 1985; Harrison et al. 1999) and an earlier thermodynamics curriculum development project (Clark 2000, 2004; Lewis 1996; Linn and 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 sufficiently elaborated to enable other students to note and discuss differences in their conceptions. Second, the pull-down menu format enables the discussion software to differentiate between students' principles so that

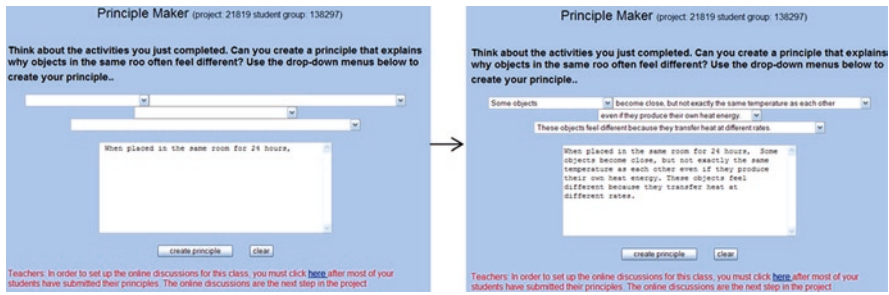


Fig. 10.6 In the PrincipleMaker explanation construction interface, students use a pull-down menu to construct an explanation from four sentence fragments that include common misconceptions

students can be automatically assigned to a discussion forum with other students who have constructed different principles to explain the same phenomenon.

Once students submit their principles, they move on to activity 7. In this activity, students participate in an asynchronous online discussion where they are encouraged to propose, support, critique, evaluate, and revise ideas. In order to foster argumentation, we designed the personally-seeded discussion software to set up and assign discussion forums to 3–5 students who have created different principles to explain the same phenomenon. This ensures that students are exposed to alternative interpretations of a given phenomenon.

10.3.3 Two Experimental Conditions

Each discussion group in the first trial of two in the original study (19 groups per condition) was randomly assigned to one of two conditions in terms of the nature of the seed comments in their discussion. The two conditions compared two seed-comment selection approaches for the discussion script. The personally-seeded groups received the explanations they constructed with the interface shown in Fig. 10.6 as their seed comments. Students in the augmented-preset groups received a pre-determined set of seed-comments constructed by the researchers using the same fragments supplied to the students in that interface. Table 10.4 shows two sets of seed comments to illustrate the difference between these two groups. The first set is the set received by the four-person augmented-preset groups. The second set is an example from a four-person group in the personally-seeded condition. The table also includes scoring information used in the original study to compare pre-post discussion gains in explanation quality.

In both conditions, the same initial scaffolding was used to enable students to explore the fragments that constituted the initial seed comments (the interface depicted in Fig. 10.6 prior to the discussions). Furthermore, the conflict schema approach was used in both conditions to for discussion groups that consisted of students with differing explanations. The two conditions diverge solely in terms of the

Table 10.4 Example sets of seed comments from discussion groups

Pre-selected Set of Seed Comments from the Augmented-Preset Groups (M=7.50, SD=3.70)		Score
When placed in the same room for 24 h, all objects become the same temperature as the room unless they produce their own heat energy. These objects feel different because they transfer heat at different rates.		12
When placed in the same room for 24 h, all objects become the same temperature as the room but only on their surface not inside them. These objects feel different because they transfer heat at different rates.		9
When placed in the same room for 24 h, objects that are good insulators stay at their original temperature regardless of the temperature of the room unless air can get inside them. These objects feel different because they are different temperature.		5
When placed in the same room for 24 h, some objects become close, but not exactly the same temperature as each other because they are made of different materials. These objects feel different because they are different temperature.		4
Example Set of Seed Comments from a Personally-Seeded Group (M=8.00, SD=2.16)		Score
When placed in the same room for 24 h, objects that are good conductors become the same temperature as the room unless they produce their own heat energy. These objects feel different because they are different temperature.		10
When placed in the same room for 24 h, hot objects become the same temperature as the room even if they produce their own heat energy. These objects feel different because they transfer heat at different rates.		9
When placed in the same room for 24 h, metal and glass objects become close, but not exactly the same temperature as the room unless they produce their own heat energy. These objects feel different because they transfer heat at different rates.		8
When placed in the same room for 24 h, all objects are at a different temperature than other objects in the same room because they are made of different materials. These objects feel different because they transfer heat at different rates.		5

third component (i.e., the nature of the initial seed comments). The augmented-preset seed comments were constructed to represent an optimized range of student misconceptions as opposed to including students' own explanations as the seed comments.

10.4 Discussion of Findings with Coding Scheme Only

Analysis of pre-post explanation gains in the original study showed that students in the augmented-preset condition demonstrated significant gains on their explanations from the first trial. A secondary analysis using the core coding scheme described earlier in this chapter and in Clark and Sampson (2008) was then conducted to provide additional insight into possible differences in the discussions in each condition that might have contributed to the observed differences in the pre-post gains in explanation quality. We now present an overview of the results from those analyses reported in the original study as a foundation for considering the potential value of using sequential analysis in tandem with the core coding scheme.

10.4.1 *Conceptual Quality*

An independent samples t -test showed that the mean conceptual quality level per episode of the comments in the augmented-preset condition ($M=1.38$, $SD=1.04$) was significantly higher than the mean in the personally-seeded condition ($M=1.21$, $SD=0.77$), $t(422)=1.94$, $p < .05$. Clark, D'Angelo, and Menekse hypothesized that this might have resulted from the fact that students in the augmented-preset condition always received at least one fully normative explanation in a seed comment. Students in the personally-seeded condition received seed comments that were based solely on their own explanations – explanations that did not necessarily include fully normative explanations.

10.4.2 *Grounds Quality and Frequency of Rebuttals*

A few other noted differences in the discussion quality between the conditions suggested certain advantages of using the augmented-preset approach. These differences were not statistically significant, but followed trends from earlier studies and thus invited speculation. The mean grounds quality level of comments in the augmented-preset condition was higher, for example, than the mean in the personally-seeded condition. The students in the augmented-preset condition thus appeared to be more likely to include grounds for their statements as opposed to focusing on connecting the statements between individual participants. Similarly, the frequency of rebuttals in the augmented-preset condition was higher than the frequency of rebuttals in the personally-seeded condition. This may have been another function of the personal connections in the sense that students were less willing to rebut or contradict an explanation when it was “owned” by another person, in comparison to when the explanation was attributed to a non-present third party.

10.4.3 *Discourse Moves*

Figure 10.7 provides an overview of the numbers and types of discourse moves made by students in each condition. The overall patterns are very similar. One difference between the two groups is that the changing of claims occurred only once in the augmented-preset group (as opposed to 13 times in the personally-seeded group) – a meta-cognitive operation that would appear to be a critical part of the learning process. One possible explanation for this finding is that the students in the preset-augmented condition did not feel that they were examining their own ideas.

Lastly, students in the personally-seeded condition contributed higher word totals and numbers of comments (although not significantly higher in the current study) than the students in augmented-preset condition. Although this difference was not as large as in our previous studies, this finding suggests that students in the personally-seeded

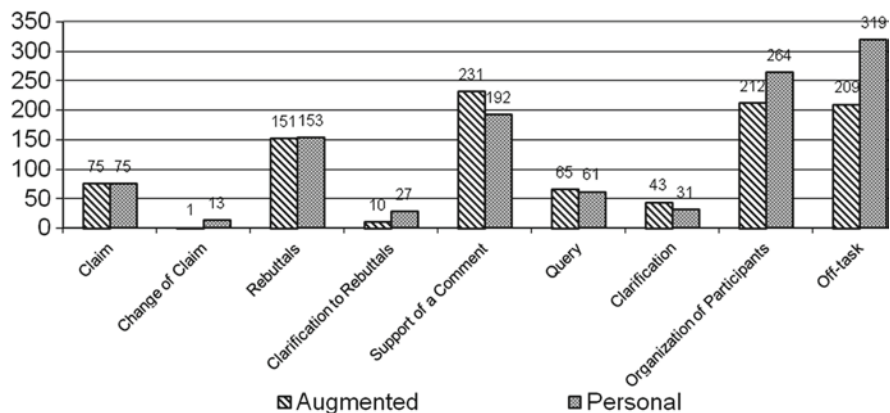


Fig. 10.7 Number and types of discourse moves in each condition

condition tend to type more than the students in the augmented-preset condition, which might suggest higher levels of engagement. As a result, the personally-seeded discussions appear to offer certain advantages as well as disadvantages when compared with the use of augmented-preset discussions along an “objectivity” *versus* “engagement” continuum.

10.4.4 Level of Opposition

Analysis of the structural level of opposition, however, showed no significant difference between the augment-preset and the personally-seeded discussion conditions in terms of the proportion of discourse episodes coded at each level of opposition, $\chi^2(5) = 2.83$, $p = .72$ (Fig. 10.8).

Another suggestive, but not statistically significant, difference between the augmented-preset and personally-seeded conditions involves the frequency of off-task comments. Approximately 28% of all comments in the personally-seeded condition were coded as off-task compared to 21% of the comments in the augmented-preset condition. This pattern was observed in a previous study (Clark et al. 2008). While both groups had many off-task comments about completely non-related topics, only the personally-seeded condition included comments that were personally focused in terms of applying social pressures to shift opinions (e.g., “we should pick mine” or “don’t pick his”).

One possible reason as to why the personally-seeded condition produced more off-task comments per student is because these students were more inclined or motivated to defend/support their own explanations and persuade others to accept their own explanations. While both groups produced off-task comments were not at all related to the topics (e.g., “nice haircut!”), only the personally-seeded condition produced comments (comment that were coded as off-task in this particular

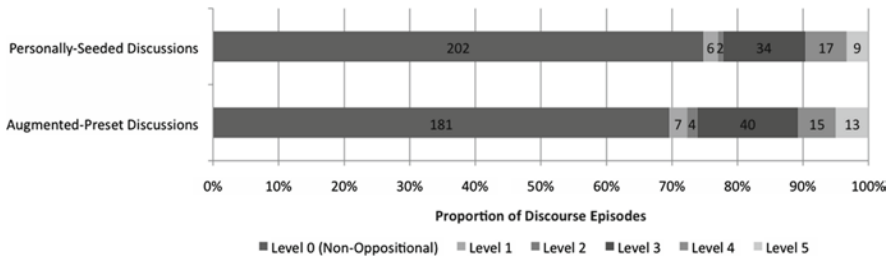


Fig. 10.8 Number of discourse episodes in the augmented-preset discussions and in the personally-seeded discussion at each level of opposition

study) that were aimed to persuade and draw group consensus. For example, students in the personally-seeded condition produced comments such as, “we should pick mine” or “don’t pick his.” These persuasive and consensus making types of comments may deserve a separate coding category in future studies.

Based on these analyses, the original study suggested that the personal embeddedness and engagement of the personally-seeded condition ultimately appeared to offer advantages as well as disadvantages compared to the augmented-preset condition along an “objectivity” *versus* “engagement” continuum. Overall, however, the core coding analysis of the ways students proposed, supported, evaluated, and revised ideas indicated that the augmented-preset condition seemed to be superior to the personally-seed discussions.

10.5 Discussion of Findings Using Sequential Analysis in Tandem with the Core Coding Scheme

10.5.1 Statistical Analysis

To sequentially analyze and identify differences in discourse patterns between conditions, the data (Fig. 10.9) used for this particular analysis consisted of 1,571 (2,160 total messages – 589 message were coded as “Other” when the categories were collapsed for the sequential analysis and thus omitted). Figures 10.9 and 10.11 present a breakdown of the observed frequencies and relative frequencies of messages (left column) and the most immediate and/or direct responses (at lag 0) to messages (top row). Messages that were posted in subsequent replies to an earlier message but separated by one or more previous responses (at lag 1 or more) were not examined in this study. The cell frequencies presented in bold identify response frequencies that were significantly higher than expected frequencies based on *z*-score tests at $p < .01$ (Fig. 10.10). The cell frequencies in italic and underlined identify response frequencies that were significantly lower than expected frequencies. The *z*-scores were computed for each possible event pairing while taking into account the differences in relative and observed frequencies of both given and target events. See Bakeman and

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp	Replies	No Replies	Givens	% replies	% givens
CLAIMa	1	88	16	0	0	109							214	1	75	.23	.05
REBUTa	0	25	14	7	0	47							93	58	151	.10	.10
QUERYa	0	7	48	0	1	15							71	29	108	.08	.07
CLARIFYa	0	2	0	1	0	1							4	5	10	.00	.01
CHANGEa	0	0	0	0	0	0							0	1	1	.00	.00
SUPPORTa	0	22	17	1	0	50							90	117	232	.10	.15
CLAIMp							1	87	12	<u>0</u>	4	86	190	0	75	.21	.05
REBUTp							0	19	19	26	4	31	99	47	153	.11	.10
QUERYp							0	4	36	0	2	<u>8</u>	50	32	92	.05	.06
CLARIFYp							0	6	3	1	0	10	20	10	27	.02	.02
CHANGEp							0	0	0	0	0	2	2	7	13	.00	.01
SUPPORTp							0	22	14	0	0	47	83	87	192	.09	.12
	1	144	95	9	1	222	1	138	84	27	10	184	916	837	1571		

Fig. 10.9 Frequency matrix from DAT with observed response frequencies to given messages. *Note:* a=message posted in augmented-preset condition; p=message posted in personal-seeded condition; bold values=higher than expected frequency; italic underlined values=lower than expected frequency

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp
CLAIMa	1.81	11.66	-1.59	-1.66	-0.55	10.41						
REBUTa	-0.34	3.12	1.56	6.75	-0.34	6.24						
QUERYa	-0.29	-1.41	16.47	-0.87	3.45	-0.64						
CLARIFYa	-0.07	1.89	-0.68	4.88	-0.07	0.04						
CHANGEa	0.00	0.00	0.00	0.00	0.00	-0.01						
SUPPORTa	-0.33	2.39	2.79	0.13	-0.33	7.30						
CLAIMp							1.96	13.30	-1.53	<u>-2.70</u>	1.51	9.73
REBUTp							-0.35	1.22	3.66	14.52	2.99	2.95
QUERYp							-0.24	-1.44	15.83	-1.27	2.04	-0.74
CLARIFYp							-0.15	1.89	0.91	0.55	-0.48	3.38
CHANGEp							-0.05	-0.60	-0.45	-0.25	-0.15	2.82
SUPPORTp							-0.32	3.06	2.55	-1.66	-1.00	8.71

Fig. 10.10 Z-score matrix revealing frequencies that were significantly higher (*bold values*) and lower (*italic, underlined*) than the expected frequency based on z-score tests at $p < .01$

Quera (1995, p.109) for more details on how the z-scores are computed in a way that takes into account the number of observed responses per category (marginal totals per column). As a result, the z-score values can be used as a means to operationally define what is to be considered (or not considered) a “discourse pattern”.

Relative frequencies were computed from the frequency matrix with DAT and reported in a transitional probability matrix (Fig. 10.11). For example, the upper

	CLAIMa	REBUTa	QUERYa	CLARIFYa	CHANGEa	SUPPORTa	CLAIMp	REBUTp	QUERYp	CLARIFYp	CHANGEp	SUPPORTp
CLAIMa	.005	.41	.07	.00	.00	.51						
REBUTa	.00	.27	.15	.08	.00	.51						
QUERYa	.00	.10	.68	.00	.01	.21						
CLARIFYa	.00	.50	.00	.25	.00	.25						
CHANGEa	.00	.00	.00	.00	.00	.00						
SUPPORTa	.00	.24	.19	.01	.00	.56						
CLAIMp							.01	.46	.06	<i>.00</i>	.02	.45
REBUTp							.00	.19	.19	.26	.04	.31
QUERYp							.00	.08	.72	.00	.04	.16
CLARIFYp							.00	.30	.15	.05	.00	.50
CHANGEp							.00	.00	.00	.00	.00	1.00
SUPPORTp							.00	.27	.17	.00	.00	.57

Fig. 10.11 Transitional probability matrix revealing probabilities that were higher (*bold values*) and lower (*italic underlined values*) than expected probabilities

left corner of the transitional probability matrix shows that 41% of all responses to claims (CLa) in the augmented-preset group were rebuttals (RBa) in contrast to 46% in the *personally-seeded group* were rebuttals (CLp → RBp). To help reveal all the differences in response patterns between the two groups, DAT translated the relative frequencies into transitional state diagrams (Fig. 10.12). The top diagram reveals discourse patterns in the augmented-preset group, and the bottom diagram reveals discourse patterns (frequencies that were higher than the expected frequency) in the personally-seeded group.

10.5.2 Differences in Transitional Probabilities

A comparison of the transitional state diagrams in Fig. 10.12 reveals the response patterns between the two groups were quite similar overall. Nevertheless, the diagrams show that students using the augmented-preset threads were more likely to post responses to claims with supporting/grounding statements (51% of responses to claims) than students that used personally seeded threads (45%). More importantly, students using the augmented-present threads were also more likely to follow up and/or respond to rebuttals with supporting statements (51%) than students using personally-seeded threads (31%). These results suggest that when students use augmented-preset threads, they are more likely to support their ideas when they respond to claims or to rebuttals of a claim. A Chi-Square test showed that the distribution of support statements elicited across the six response categories were significantly different between the two groups $\chi^2(5) = 14.1, p = .015$. These particular findings help to illuminate when and where students tend to support their ideas. The implications of this finding is that if students are encouraged and/or provided

additional guidance on how to produce a greater number of rebuttals to each claim, we can expect to see the number of support statement increase as well.

10.5.3 Differences in the Mean Number of Responses Elicited Per Message

We conducted a 2 (conditions)×4 (type of *oppositional* exchange) ANOVA to test for differences in the frequencies of four types of *oppositional* message-response exchanges – exchanges where rebuttals were posted in reply to claims, and where *oppositional* comments were posted in reply to rebuttals (i.e., claim → rebuttal, rebuttal → rebuttal, rebuttal → query meaning, rebuttal → change claim). We chose to select these four *oppositional* exchanges for this analysis based on the assumption that deeper inquiry is driven by the juxtaposition of differing viewpoints and to help reduce the chances of committing Type I error. We also conducted a 2 (conditions)×6 (*supportive comments* posted in reply to claims, rebuttals, queries, clarifications, change claims, and *supportive* comments) ANOVA to test for differences across the six primary types of *supportive* exchanges (i.e., claim → supporting comment, rebut → support, query → support, clarify rebuttal → support, change claim → support, support → support) based on the differences noted in the transitional state diagrams.

10.5.3.1 Oppositional Exchanges

No significant differences were found between conditions in the number of responses posted across the four types of *oppositional* exchanges, $F(1, 1,054) = .00$, $p = .982$. The results of the sequential analysis revealed no indications that one condition lead to higher levels of argumentation in terms of the *oppositional* exchanges than the other condition. Significant differences were found in the number of responses elicited per message between the four different types of exchanges, however, independent of condition, $F(3, 1,054) = 211.26$, $p = .000$. In other words, certain types of exchanges tended to elicit more responses than other types of exchanges. Claims elicited on average 1.17 rebuttals ($STD = 1.06$, $n = 150$), rebuttals elicited .14 counter rebuttals ($STD = .396$, $n = 304$), .11 queries ($STD = .332$, $n = 304$), and .01 change claims ($STD = .114$, $n = 304$). No interaction was found between *oppositional* exchange type and condition, $F(3, 1,054) = .36$, $p = .780$.

10.5.3.2 Supportive Exchanges

No significant differences were found between conditions in the number of responses posted across the six *supportive* exchanges examined in this study, $F(1, 1,117) = .00$, $p = .983$. We found no indication that one condition lead to

higher levels of supportive exchanges (as opposed to argumentative exchanges) than the other condition. Significant differences were found in the number of responses elicited per message between the six supportive exchanges independent of condition, $F(5, 1,117)=90.86, p = .000$. The average number of supporting comments posted in reply to claims was 1.30 ($STD=1.11, n=75$ claims), .26 for rebuttals ($STD = .52, n=304$), .12 for queries ($STD = .22, n=200$), .30 for clarify rebuttals ($STD = .46, n=37$), .14 for change claims ($STD = .36, n=14$), and .23 for supporting comment ($STD = .47, n=424$).

We did find an interaction between type of supportive exchanges and conditions (Table 10.5), $F(5, 1,117)=2.42, p = .034$. In the augmented-preset group, students posted 27% more supportive comments in reply to claims, 54% more in reply to rebuttals, and 60% to queries. In contrast, students in the personally-seeded group posted 270% more supportive comments in reply to clarifying rebuttals, and 14% more in reply to supportive comments. We conducted additional analysis and found that: (a) claims, rebuttals, and queries (messages that elicited more supporting comments in the augmented-preset group than the personally-seeded group) were posted on average at 3.00 ($STD=2.07, n=654$) thread levels deep in discussion threads; (b) clarify rebuttals, change claims, and supportive comments (messages that elicited more supporting comments in the personally-seeded group) were posted on average at 3.23 ($STD=1.75, n=475$) levels deep; and (c) that this observed difference in thread level was statistically

Table 10.5 Mean number of supportive comments posted in reply to discourse moves between groups

Exchange	Group	Mean	std	n
Claim-support	Preset	1.45	1.26	75
	Personal	1.15	.93	75
	Total	1.30	1.11	150
Rebut-support	Preset	.31	.57	151
	Personal	.20	.46	153
	Total	.26	.52	304
Query-support	Preset	.14	.35	108
	Personal	.09	.28	92
	Total	.12	.32	200
ClarifyRebuttal-support	Preset	.10	.32	10
	Personal	.37	.49	27
	Total	.30	.46	37
Change-support	Preset	.00		1
	Personal	.15	.38	13
	Total	.14	.36	14
Support-support	Preset	.22	.46	232
	Personal	.24	.48	192
	Total	.23	.47	424
Total	Preset	.38	.76	577
	Personal	.33	.62	552
	Total	.36	.69	1,129

significant, $t(1,127) = -1.96$, $p = .05$. This finding suggests that students in the augmented-preset threads tended to reply with supportive comments to comment types that occurred earlier in a discussion thread, where as students in the personally-seeded discussions tended to reply with supportive comments to comment types that occurred later in a discussion thread.

10.6 Affordances of Using Both Methods in Tandem

In summary, the analysis performed with the core coding scheme alone suggested that (a) augmented-preset threads produced comments with higher conceptual quality (or more normative explanations as defined in Clark and Sampson 2008); (b) augmented-preset threads *may* have helped to produce more grounded claims; (c) augmented-preset threads *may* have helped to produce more rebuttals on the grounds of each claim; and (d) no differences were found in the proportion of episodes across each of the five levels of opposition (Table 10.6). The sequential analysis then revealed that (a) the patterns of discourse between the groups were overall very similar in structure; (b) there were no significant differences in the number of oppositional exchanges produced by students between the groups; (c) the number of responses posted in reply to each message depended heavily on the function or type of message; and (d) the time and place where students reply with supporting comments depends both on the type of message they are replying to and whether students are using augmented-preset *versus* personally-seeded discussions. Students using augmented-preset threads, in other words, were more likely than students using the personally-seeded threads to respond to claims and rebuttals with supporting statements. In this case, using both methods in tandem enabled us to: (a) pinpoint where, when, why, and/or how particular types of discourse moves of interest are elicited within the course of a conversation; and (b) identify where and how changes in the discourse process can be made to help increase the frequency of discourse moves of particular interest. In all, the sequential analysis revealed patterns that were more or less consistent with the previous findings reported by Clark and Sampson and also provided a quantitative and process-oriented approach to describing the nature and quality of argumentation in these different discussion forums (see Table 10.6).

Overall, our sequential analysis of the data sequential analysis produced potential explanations for the earlier findings identified with the core coding scheme and provided further insights into the patterns that emerged within the students' discourse. We therefore believe that using the core coding scheme in tandem with sequential analysis can provide useful insights into online discourse by providing visual representations (including quantitative measures) of the discourse process in ways that can help use better understand of how online discussion forums (both asynchronous and synchronous) affect the way discourse unfolds over time (when examined at the micro level) and how changes in processes help to produce quality argumentation.

Table 10.6 Findings on the effects of using augmented-preset discussion threads using core coding scheme and sequential analysis

Outcome	Core coding scheme	Sequential analysis
Conceptual quality	Higher	*Not investigated*
Level of grounding	Higher (but NS)	*Not investigated*
Rebuttal frequencies	More rebuttals (but NS)	State diagrams revealed no major differences in patterns (transitional probabilities with significant z-scores) in the messages that preceded and followed each rebuttal.
Level of opposition	No difference in proportion of episodes across each level of opposition (based on holistic evaluation of entire discussion thread)	<p>Oppositional exchanges – No difference in mean number of oppositional responses to claims and rebuttals.</p> <p>Common patterns – Most response patterns identified from the sequential analysis were shared between groups. These patterns included the following exchanges:</p> <p>CLAIM – REBUT CLAIM – SUPPORT REBUT – REBUT REBUT – QUERY REBUT – SUPPORT QUERY – QUERY SUPPORTING – SUPPORT</p> <p>Unique patterns – Analysis of response patterns reveal that students in the augmented-preset group tended to respond back with further clarifications of rebuttals, where as in the personally-seeded group tended to respond to the clarification of rebuttals with supporting comments,</p> <p>Supportive exchanges – Differences in the way students posted supportive comments between the two groups depended on the function or discourse move that triggered students’ responses.</p>

10.7 Directions for Future Research

In future studies, we intend to examine (a) how conceptual quality correlates with specific discourse patterns, and (b) how and to what extent specific patterns help promote and/or explain observed differences in conceptual quality. We plan to explore, for example, the extent to which high *versus* low levels of oppositional

exchanges trigger/elicit subsequent comments that are more normative and/or nuanced. Given that there were few differences in response patterns observed between the two groups, a more detailed analysis of the response patterns produced by student groups *within* the augmented-preset condition might help shed light on discourse patterns that promote conceptual quality. For example, a Markov analysis can be applied to our data to determine if there are significant differences in the frequency of particular three-event chains of discourse moves (as opposed to two-event chains) that distinguish one group from the other – Markov chains that might help to explain observed differences in the quality of the group performance overall.

Our future work will also explore the relationship between response patterns and the level of grounds students include in their arguments. To examine this relationship, the Clark and Sampson coding scheme will be expanded to differentiate responses that clarify, request clarification, and support in terms of whether they focus on the grounds or the thesis of the parent comment. At present, the Clark and Sampson coding scheme only differentiates rebuttals in terms of whether or not they focus on the thesis or the grounds of the parent comment. With an elaborated coding scheme that makes this differentiation for other comment types, we will apply sequential analysis to determine the percentage of responses to comments that focus on the thesis *versus* the grounds of a claim. Next, we can determine to what extent the ratio of focus on thesis *versus* grounds affects the level of grounds observed across all messages posted within a discussion thread (and perhaps across both experimental groups). Given the assumption that students are working under limited time and resources, one can test the claim that these two goals of focusing on the thesis and the grounds of comments are working in competition or synergy with one another. Any observed tendencies in students' responses that pursue one particular goal may have an adverse effect on the extent to which they are able to accomplish other goals. The observed response tendencies can then be compared between conditions to explain any observed differences in grounding.

One potential constraint with sequential analysis (when used to examine adjacent message and responses to messages), however, is that each observed response must be explicitly mapped or threaded to the correct message stated previously within a conversational thread. Students often post responses that perform multiple discourse moves that address multiple comments from multiple messages (i.e., messages posted immediately prior to the response and posted earlier in the message thread). One way to address this limitation is to modify both the coding scheme and coding procedures. For example, we can: (a) expand the coding scheme by assigning one code for 'a rebuttal against the thesis of a *claim*' and another code for 'a rebuttal against the thesis of a *rebuttal*'; and (b) parse messages that perform multiple discourse moves into separate units and assign individual codes to each unit. Another alternative is to integrate pre-specified prompts into the discussion board to constrain each posting to respond only to the parent message while using one and only one discourse move. Although each of these solutions presents its own set of limitations or issues, these types of changes can potentially increase the accuracy and precision of the state diagrams resulting from a sequential analysis of the

students' conversations. This increased accuracy and precision would support a more detailed examination of the relationships between discourse processes, conceptual quality, and level of grounding.

One limitation of the DAT software is that the number of discourse moves presented in each state diagram is limited to a maximum of six discourse moves. The software tool will require further changes so that it can generate state diagrams which can convey transitional probabilities between larger numbers of discourse moves to conduct some of the future studies described above. Furthermore, the transitional diagrams generated with DAT will need to also convey the probabilities in which each message elicits no response. In doing so, the observed transitional probabilities (response patterns) might provide more accurate explanations for the observed differences in mean response scores (e.g. the average number of challenges posted in reply to a claim). The software will also need to include a mechanism that enables the viewer to: (a) flip and superimpose one state diagram over another diagram to make it easier to visualize and identify the similarities and differences between diagrams (particularly with diagrams containing large numbers of discourse moves); and (b) aggregate the diagrams into one diagram to reveal the similarities and differences (using links with varied colors and/or gray scale) with respect to or relative to one selected diagram. Tools for aggregating data across matrices and superimposing transitional state diagrams over another diagram can be found in the software application called jMAP (Jeong 2008). Tools like this could be integrated into DAT to facilitate the comparison of larger and more complex state diagrams.

We would also make the following additional recommendations for future research: (a) expand the analysis to measure the frequency of three-event sequences to determine whether some event pairs are more effective in eliciting desired responses than other event pairs; (b) analyze the discourse between experts/teachers and identify sequences that distinguish experts from novices using multidimensional scaling; (c) test and validate process models across variants of the task using new message codes and labels to facilitate discussions and to identify new patterns of interaction that support group performance; (d) examine how specific scaffolds and instructional strategies affect the way discourse patterns change over time (learning trajectories) by visually flipping and superimposing state diagrams of discourse patterns observed across different time periods over a target state diagrams depicting discourse patterns exhibited in the discourse between experts and teachers; and (e) assess scientific explanations by examining students' causal loop diagrams and use tools like jMAP and DAT to examine how discourse patterns trigger changes in students' causal diagrams/understanding that converge toward expert diagrams/understanding.

This chapter, overall, demonstrates the application of Clark and Sampson's coding scheme for capturing the processes of scientific argumentation and how sequential analysis can be used in tandem as a way to provide both quantitative and qualitative descriptions of discourse processes in instructional contexts. We hope that the ideas presented here will form the basis of a new process-oriented framework for measuring discourse and argumentation in CSCL environments and

developing new process-oriented methods to support, monitor, evaluate, and improve student learning and performance.

References

- Bakeman, R., & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge: Cambridge University Press.
- Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ* (p. 109). Cambridge University Press: Cambridge.
- Clark, D. B. (2000). Scaffolding knowledge integration through curricular depth. Unpublished doctoral dissertation, University of California, Berkeley.
- Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, 24(4), 467–563.
- Clark, D. B., D'Angelo, C. M., & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology*, 18(4), 321–333.
- Clark, D. B., & Linn, M. C. (2003). Scaffolding knowledge integration through curricular depth. *Journal of the Learning Sciences*, 12(4), 451–494.
- Clark, D. B. (2004). Hands-on investigation in Internet environments: Teaching thermal equilibrium. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 175–200). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, 29(3), 253–277.
- Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, 19(3), 343–374.
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45(3), 293–321.
- Clark, D. B., Schleigh, S. P., Menekse, M., D'Angelo, C. M., & Sampson, S. (2008). Improving the quality of student argumentation through the initial structuring of online discussions. Paper presented as part of symposium organized by William Sandoval for the American Educational Research Association (AERA) 2008 meeting. New York.
- Clough, E. E., & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: Bringing together scientific and personal views. *Physical Educator*, 20, 176–182.
- Cuthbert, A. J., Clark, D. B., & Linn, M. C. (2002). WISE learning communities: Design considerations. In K. A. Renninger & W. Shumar (Eds.), *Building virtual communities: Learning and change in cyberspace* (pp. 215–246). Cambridge: Cambridge University Press.
- 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(1), 63–103.
- 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.
- Erkens, G., Kanselaar, G., Prangma, M., & Jaspers, J. (2003). Computer support for collaborative and argumentative writing. In E. D. Corte, L. Verschaffel, N. Entwistle, & Jv Merriënboer (Eds.), *Powerful learning environments: Unravelling basic components and dimensions* (pp. 157–176). Amsterdam: Pergamon, Elsevier Science.
- 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(1), 55–87.
- Janssen, J., Erkens, G., Jaspers, J., & Kanselaar, G. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49, 1037–1065.
- Jeong, A. (2004). Discussion analysis tools. Retrieved March 10 2009. <http://mailer.fsu.edu/~ajeong/DAT>
- Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education*, 26(3), 367–383. Retrieved December 2008 http://garnet.fsu.edu/~ajeong/research/Jeong2005_AnalyzingInteraction.pdf.
- Jeong, A. (2007). The effects of intellectual openness and gender on critical thinking processes in computer-supported collaborative argumentation. *Journal of Distance Education*, 22(1), 1–18.
- Jeong, A. (2008). jMAP. Retrieved March 10 2009. <http://jmap.wikispaces.com>
- 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 Academic.
- Lewis, E. L. (1996). Conceptual change among middle school students studying elementary thermodynamics. *Journal of Science Education and Technology*, 5(1), 3–31.
- Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*. Mahwah: Lawrence Erlbaum.
- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education*, 12, 8–22. Retrieved July 09, 2005, from http://aied.inf.ed.ac.uk/members01/archive/vol_12/rourke/full.html.
- Schwarz, B. B. & Glassner, A. (in press). The role of CSCL argumentative environments for broadening and deepening understanding of the space of debate. In R. Saljo (Ed.), *Information technologies and transformation of knowledge*.