The Combined Effects of Response Time and Message Content on Growth Patterns of Discussion Threads in Computer-Supported Collaborative Argumentation

Allan Jeong

Abstract

This study examined the effects of response time and message content on the growth patterns of discussion threads in computer-supported collaborative argumentation. Event sequence analysis was used to measure response times between threaded messages and responses containing arguments, evidence, critiques, evaluations, and other comments from online debates. The results supported and contradicted the findings of Hewitt and Teplovs (1999). Response rates overall declined at a rate of 17% per day in wait time across all message categories. On the other hand, the posting of critiques and particular types of argumentative exchanges produced higher response rates of .72 and higher, and their average wait times of 1.04 days were significantly longer than those of other message types. The debate format and use of message labels may have produced sufficient argumentative exchanges to produce high response rates despite the long response times, which in turn helped sustain the growth of discussion threads.

Résumé

L'étude examine les effets du temps de réponse et du contenu du message sur les modèles de croissance du volume des discussions dans l'argumentation collaborative assistée par ordinateur. L'analyse séquentielle des événements a été utilisée pour mesurer les temps de réponse, i.e. le temps entre les messages envoyés et les réponses reçues qui contenaient des arguments, des preuves, des critiques, des évaluations et d'autres commentaires issus des débats en ligne. Les résultats ont appuyé et contredit les résultats de Hewitt et Teplovs (1999). D'une part, le rythme des réponses a diminué dans l'ensemble, à un taux de 17 % par jour en temps d'attente dans toutes les catégories de message. D'autre part, les articles de critique et d'argumentation ont produit des taux de réponse plus élevés, soit de 72 % et plus, et leur temps d'attente moyen, 1,04 jours, était nettement plus long que celui des autres types de message. Il est possible que la formule débat et l'utilisation de labels pour identifier les messages aient produit suffisamment d'échanges d'argumentation entraînant des taux de réponse élevés, malgré les longs temps de réponse, qui à leur tour ont favorisé la croissance du volume des discussions.

Introduction

Computer-mediated communication (CMC) is widely used in distance learning (Harasim, 1993; Berge, 1997) to support student-student interaction and online group discussions. The threaded discussion board is one application of CMC that allows students to participate asynchronously, posting messages to group discussions at any time and from anywhere. The ability to participate in discussions independent of time is one of the recognized advantages of asynchronous discussions (Harasim, 1993) because students can read, reflect, compose, and post responses at their own time and convenience. Students can revisit, build on, and respond to messages to advance discussions in earlier threads and not just in the most current threads. In other words, multiple discussion threads can develop concurrently over time. As a result, CMC has been found to generate higher levels of critical thinking and ideas that are more important, justified, or linked than face-to-face or synchronous discussions (Webb, Newman, & Cochrane, 1994; Newman, Johnson, Cochrane, & Webb, 1996). At the same time, asynchronous discussions can also have potential disadvantages because students must wait for replies over the course of several hours or days, which can hinder the momentum and flow of discussions. Students also face the cognitive challenge of monitoring and contributing to multiple and concurrent discussion threads.

Hewitt and Teplovs (1999) found that discussion threads were more likely to maintain activity and growth when replies were posted within a day of the most recent posting. Based on the analysis of 4,086 messages posted in an asynchronous threaded discussion environment from seven graduate-level distance education courses, the findings showed that responses posted to a thread within 24 hours resulted in a .26 to .63 probability of eliciting additional responses depending on the number of existing messages in a thread. After a day of inactivity, the odds of maintaining an active thread dropped to .18 to .41. After two days of inactivity, the odds dropped further to .12 to .31. The steady decrease in activity in discussion threads was attributed to increased competition for students' attention and responses from other discussion threads that grew in number with each passing day (Hewitt, 2003). As new threads compete for students' responses, messages in earlier threads are less likely to receive the responses needed to keep the threads growing. These findings suggest that discussion threads can be negatively effected by long response times commonly observed in asynchronous discussions. Hewitt recognized that the content or function of messages must also be taken into consideration in determining how likely it is that a discussion thread can maintain the interest and attention of its readers and consequently its continued growth over time. However, no studies at this time have examined the combined

effect of message content and response time on the growth of discussion threads.

The lack of studies on the effects of message content and response time can be attributed to the difficulties of analyzing the content of computer conference messages (Rourke, Anderson, Garrison, & Archer, 2001). The major difficulty in analyzing message content is establishing the unit of analysis. Computer conferencing messages more often than not contain multiple ideas that perform multiple functions. As a result, the contents of a message must be classified into multiple codes, making it difficult if not impossible to map message-response sequences in terms of predefined message categories (Levin, Kim, & Riel, 1990; Newman et al., 1996; Gunawardena, Lowe, & Anderson, 1997). The problems with mapping message-response sequences have deterred researchers from examining differences in response times between messages and responses with respect to their content. Consequently, it has not been possible to measure the effect of message content and their associated response times on the growth of discussion threads.

One approach to solving the problems in mapping message-response sequences is to require students to preclassify their contributions to discussions using a predetermined set of message/response categories. This constrains each message to serve only one function at a time and establishes the message as the unit of analysis. This approach has been used and evaluated in a number of computer-supported collaborative argumentation (CSCA) systems. The ACT system (Sloffer, Dueber, & Duffy, 1999), for example, is a threaded discussion board that is designed to scaffold online debates by requiring students to preclassify each posting to one of six response categories: proposal, counter-proposal, supporting reasons, detracting reasons, supporting evidence, and detracting evidence. Mc-Alister (2003) proposed a synchronous chat tool to support collaborative argumentation by requiring students to preclassify messages to inform, question, challenge, reason, support, or maintain chat discussions. In each of these response categories, students are able to choose a specific sentence opener (e.g., "A counterargument is …") to channel students' thoughts by the process of completing the sentence in a way that fits with the opener. Belvedere (Suthers, 1998) is a concept-mapping tool to support argumentation by constraining responses to hypotheses, claims, data, principles, and backings. Students also identify the relationships between statements with links to express support or opposition.

This study used the same techniques observed in CSCA to classify each message/response into response categories designed to facilitate collaborative argumentation in asynchronous threaded discussions. With each message/response clearly classified and identified by response category, the sequence of message and responses could be examined. As a

result, the response times between specific types of messages and responses could be measured to determine the combined effects of response time and message content on the growth patterns of discussion threads. Using the constraint-based approach to scaffold argumentation in a threaded discussion board, the content of messages in five online debates was examined in terms of cognitive events that support critical thinking (Gunawardena et al., 1997; Garrison, 1992) and argumentation (Cerbin, 1988). These cognitive events included the statement of arguments, supporting evidence, critiques, elaboration, evaluation, and process comments. To measure how response time and message content affected the growth of discussion threads, event sequence analysis (Bakeman & Quera, 1995; Bakeman & Gottman, 1997) was used to examine the relationship between response rates and response times between specific types of messages and message-response sequences. Using this combination of methods, the following questions were addressed in this study.

1. *Response rates.* For each particular type of message, what percentage of the messages (lag 0 events) elicited responses (lag 1 events)? In other words, what is the response rate for each type of message? Is a critique or disagreement more likely to elicit a response than an agreement or elaboration?

2. *Effects of message content on the wait time for responses.* Do particular types of messages generate longer wait times between the time a message (lag 0 event) is posted and the time a reply to the message (lag 1 event) is posted? For example, is the wait time for a response to arguments shorter than the wait time for responses to critiques?

3. *Effects of response time on response rates.* Does the amount of elapsed time between the time of posting of a specific type of message (lag 1 event) and the time of posting of a response to the message (lag 0 event) affect the ability of the message (lag 1 event) to elicit subsequent responses (lag 2, lag 3, and onward)? In other words, how does response time in posting a particular type of message affect its response rate and the number of subsequent responses?

Method

Participants

The participants were 19 graduate students from a major university in the Southeast region of the United States consisting of eight women and 11 men ranging in age from 24 to 49. The students who willingly agreed to participate in this study were enrolled in an online course on theories of learning and cognition.

Online Debates

The group discussions examined in this study were collected from five weekly debates that examined the assumptions and grounds of various learning theories. For example, two of the discussion topics were "Knowledge cannot be instructed (transmitted) by a teacher—it can only be constructed by the learner" and "Schema theory is more of a constructivist theory than a cognitive theory." The debates were conducted on threaded discussion boards using the Blackboard system.

For each debate, students were randomly assigned to one of two opposing teams. The debate teams were balanced by gender and level of participation observed in previous discussions in the course. One team was assigned to support a given claim or position, and the other was assigned to challenge the position. Students were instructed only to post messages that supported their team position by posting arguments, supporting evidence, and by challenging the messages posted by the opposing team. Students were required to post a minimum of four messages in each debate. After each debate students were also required to vote on the team that won the debate based on the team that presented the most convincing arguments. Twenty-five percent of the course grade was awarded for participation in the debates and other discussions throughout the course.

Students were required to preclassify their messages by message category (see Table 1) when they posted messages to the discussion board. For each category students inserted designated labels into the subject headings of each message posted to the discussions. The labels consisted of eight message categories that included position statements, arguments to support their assigned position, evidence to support stated arguments, criticisms to challenge opponent's statements, elaboration, judgments for drawing conclusions, evaluation of comments, and other process comments. Each label had to be followed by an additional tag, o=opposing team or s=supporting team, to identify team membership (see Figure 1). These procedures were implemented in order to ensure that each message addressed only one function at a time, because more often than not students compose messages that address more than one function or topic of discussion at the same time. This enabled each message to be examined as a unit of analysis. The effects of constraining students' postings to specific categories to scaffold student discussions (Duffy, Dueber, & Hawley, 1998; Sloffer et al., 1999) are currently under investigation.

Data Analysis

Messages and response codings. The five debates generated a total of 565 messages. To establish the accuracy of students' message labels used to classify each message by response category, inter-rater reliability was

Table 1
Coding Scheme for Student Labeled Messages in the Online Debates

Label*	Category	Definitions
POS	Position statement	State your position on the issue
ARG	Arguments	Establishing arguments to support or oppose a given issue or position
EVID	Evidence	Providing evidence and examples to support a stated argument
CRIT	Critique	Examine and find flaws or weakness in another's response
ELAB	Elaborate	Expanding on an idea provided by another
EVAL	Evaluation	Analyze and determine value of a response
JUDG	Judgment	State a judgment or position on a stated argument based on presented evidence and analysis
OTH	Other	Process comments or extraneous comments not relevant to the debate.

*Each label was followed by a tag *o* to identify postings from the opposing team, and a tag *s* to identify postings from the supporting team.

computed by comparing the student's self-codings with the codings of the experimenter. The inter-rater reliability was good with Cohen's Kappa 0.68 (Bakeman & Gottman, 1997). All messages that were posted as a reply and were threaded to a previous message were identified by rule as a response.

Data analysis. Two computer programs, Forum Manager and Discussion Analysis Tool (DAT), were developed to perform the data analysis. Forum Manager was used to download threaded discussions from Blackboard into Microsoft Excel, compile the message labels and response time data, and identify the message-response sequences conveyed in the hierarchical organization of the threaded messages. DAT was then used to compute the frequencies of each message by category and the frequencies of specific responses to each message, convert the response frequencies into relative frequencies (or transitional probabilities), generate a visual representation of the response probabilities in a transitional state diagram, compute the response times and wait times between each message-response exchange, and to compute the Z-scores for each message-response exchange to determine which exchanges (or event sequences) were prevalent patterns of interaction that occurred at rates significantly higher than the expected probability (Bakeman & Gottman, 1997).

The eight original message categories were reduced to six categories to minimize the total number of possible event sequences and to achieve sufficient cell frequencies for measuring the probabilities of each observed message-response pairing. Messages labeled as Position Statements were

1 **ARGo:** To say that 'knowledge cannot be instructed or transmitted by a teacher, but can only be constructed by the learner' is an overgeneralization. It would make teachers obsolete.

2.   **ARGo: Excellent argument. It is the symbiotic relationship between teacher and student that enables learning to take place. The teacher must provide guidance to the learner.**

3 . . .   **JUDGs:** This is a well-reasoned commentary. I agree with these contentions, but I believe that what has not been explained is how a learner constructs new ideas stemming from new theories or perspectives. Although I side with the opposition, I believe there must be a middle ground that recognizes that learners can construct knowledge.

4 . . . . .   **ELABo:** A self-taught person can only construct new ideas until they have learned the pre-requisite information and skills needed to construct the new knowledge.

5 . . . .   **OTHo:** Very well stated about the conditions of self-taught individuals. I thought that was a good comeback.

6   **CRITs:** What teachers transmit is not knowledge, but accumulated information. We use this information as tools to construct new knowledge. By analogy, we need access to tools and materials to construct a house. The transmission and access to tools and information simply facilitates the construction of a house.

7 . . .   **CRITo:** But I would argue that teachers can and do transmit knowledge or information. What you are saying instead is that NEW knowledge (or building an architecturally unique house) cannot be instructed by a teacher and must be constructed by the learner. And that is an entirely different argument.

8 . . . . .   **CRITs:** But not all learned knowledge is instructed. I learned how to build houses by mainly watching my father build houses and learned through practice. I never received any formal instruction.

9 . . . .   **EVALs:** From what has been said thus far, some learning result from instruction and some do not. From an epistemological standpoint, this is consistent with Constructivists' acceptance and synthesis of cognitive learning theories and principles.

10 . . .   **ARGo:** Although there is truth to what you say, most of the skills and tools used to build a house is passed down from one generation to the next through and are not re-constructed. In other words, we generally try to avoid the mistake of always trying to "re-create the wheel."

11 . . . . .   **EVIDo:** Your ideas are consistent with what Driscoll states in Chapter 3 of the textbook. Teacher input is important and so is their assessment of the learner's performance in order to ensure performance is to established standards.

The texts in the example have been modified and abbreviated for illustrative purposes.

Figure 1. Excerpt from a discussion thread with student-labeled messages.

aggregated with the Arguments category because students presented their position statements with supporting arguments. Furthermore, students were already preassigned to a given position before each debate. Judgment was collapsed with Evaluations because Judgments were difficult to discriminate from evaluation statements. Few Judgments were observed in the debates because students were not expected or instructed to make

summary judgments of the presented arguments. Finally, the tags used to identify team membership were removed from the message labels in order to analyze the collective interactions between the supporting and opposing debate teams.

Theoretical assumptions. The theoretical basis for examining group interactions in terms of response times and event sequences observed in threaded message-response exchanges are based on the assumptions of dialogic theory (Bakhtin, 1981; Koschmann, 1999). The theory views language as part of a larger whole or social context in which all possible meanings of a word interact, possibly conflict, and affect future meanings. Meaning is produced not by examining an utterance by itself, but by examining the relationship between utterances. Meaning is therefore renegotiated and reconstructed as a result of conflict in social interactions, which drives inquiry, reflection, articulation of individual viewpoints, and underlying assumptions. Given these assumptions, the focus of this study was to analyze the relationship between threaded messages (or message-response exchanges) with respect to the content of the messages and responses and the response times separating the messages and responses. Of most interest in this study were the types of exchanges that drew out conflict and disagreement based on the assumption that conflict drives inquiry and dialogue and consequently the growth of discussion threads.

Discussion of Findings

Response rates. Table 2 displays data to describe the various relationships between messages and responses in students' exchanges in the threaded discussions and online debates. The table shows, for example, that 150 arguments were posted in the debates, and 49 of these 150 arguments did not elicit a response. Given that 101 of the 150 arguments were successful in eliciting one or more responses, the *response rate* for arguments was .67. To understand better how arguments generated such a response rate, Table 2 shows the distribution of responses to arguments across the six message categories. Among the 145 messages that were posted in response to arguments, .26 were follow-up arguments, .19 were supporting evidence, .28 were criticisms, and only .09 were evaluations of the arguments. The messages that generated the highest response rates and were most likely to contribute to the growth of discussion threads were critiques (.72) and arguments (.67). The messages that were least likely to contribute to the growth of discussion threads were evidence (.54), evaluative responses (.54), and *other* comments (.40). The high response rate to critiques is consistent with the assumptions of dialogic theory: this conflict in message-response exchanges drives the processes of inquiry and critical discourse.

Table 2
Transitional Probabilities Between Message-Response Interactions

	ARG	EVID	CRIT	ELAB	EVAL	OTH	Replies	No Replies	Givens	Reply Rate	Wait Time	Response Time
ARG	**.26**	**.19**	**.28**	**_.12_**	**_.09_**	**_.06_**	145	49	150	.67	.99	.94
EVID	.13	.13	.29	.13	.16	.16	62	35	76	.54	.96	.98
CRIT	.19	.17	.23	.15	.14	.13	88	25	89	.72	1.04	.96
ELAB	**_.10_**	.08	**_.10_**	**.35**	.21	.17	78	30	85	.65	.68	.96
EVAK	.16	**_.05_**	**_.02_**	.20	**.29**	**.29**	56	39	85	.54	.74	.74
OTH	.14	**_.00_**	**_.06_**	**_.06_**	.20	**.54**	35	48	80	.40	.50	.67
	84	60	89	79	74	78	464	226	565	.59	.82	.88

The probabilities displayed in boldface are higher than expected probability, and the probabilities in boldface/underscore are lower than expected probability based on Z-score tests in Table 3.

The overall response rate to messages was .59, given that a total of 226 messages received no replies of the total of 565 posted messages. This overall response rate was high relative to Hewitt and Teplovs' (1999) finding where threads varied from a .26 to .63 probability of remaining active when responses were received within a day. The high response rates observed in this study can be attributed to the use of a debate format that required the opposing teams to critique arguments and post rebuttals. Something that may have also contributed to the high response rates was the use of message labels, which enabled students to locate arguments and critiques quickly among multiple discussion threads. In turn, this may have enabled students to respond and maintain more active discussion threads. In Figure 1, for example, the label CRITs in message 6 may have helped students on the opposing team to identify, evaluate, and respond to criticisms from the supporting team. Similarly, the labels may also have helped students on the opposing team to locate opposing arguments (message 10) posted by fellow team members and respond with evidence to support the argument (message 11).

The message-response exchanges that generated conflict and disagreement were found to be some of the most prevalent patterns in students' interactions. The Z-scores displayed in Table 3 show that critical responses to arguments (Z-score=3.10, alpha=.05, n=40) and evidence (Z-score=2.12, alpha=.05, n=18) were significantly higher in frequency than the expected frequency based on random chance alone. Also prevalent in the observed interactions were the responses to arguments with follow-up arguments

Table 3
Z-Scores to Identify Transitional Probabilities that Deviate
From Expected Frequencies

	ARG	EVID	CRIT	ELAB	EVAL	OTH	
ARG	**2.80**	**2.76**	**3.10**	**−1.78**	**−2.77**	**−4.12**	145
EVID	−1.14	−0.01	**2.12**	−0.93	0.04	−0.15	62
CRIT	0.33	1.28	0.94	−0.62	−0.66	−1.20	88
ELAB	**−1.97**	−1.51	**−2.19**	**4.53**	1.21	−0.04	78
EVAL	−0.42	**−1.80**	**−3.53**	0.56	**2.75**	**2.51**	56
OTH	−0.61	**−2.37**	**−2.10**	**−1.35**	0.68	**6.17**	35
	84	60	89	79	74	78	464

The *Z*-scores in boldface greater than 1.95 indicate the transitional probabilities that were significantly higher than the expected frequency by chance alone. *Z*-scores that are in boldface and underscored less than −1.95 indicate transitional probabilities that were significantly lower than the expected frequency.

(*Z*-score=2.80, alpha=.05, *n*=38) and responses to arguments with evidence (*Z*-score=2.76, alpha=.05, *n*=28). Although the frequency of these patterns was found to be statistically significant, the results must be interpreted as exploratory in nature given the high probability of Type II error among the 36 possible *Z*-score tests.

Effects of message content on the wait time for responses. This study found that critiques had on average a significantly longer wait time (1.04 days) for responses than any other type of message, yet at the same time, critiques generated the highest response rate (.72) (see Table 2). This finding contradicts Hewitt and Teplovs (1999) conclusion that the growth of discussion threads, or response rates, depends on shorter wait times separating the time of posting of a message and replies to the message. Nevertheless, the finding in this study is consistent with findings from studies on the effects of extended wait time in face-to-face discussions (Tobin, 1987). The use of extended wait time in face-to-face discussions has been found to result in less failure to respond in addition to fewer low level questions, longer student responses, increases in alternative responses, increase in the complexity and cognitive level of student responses, and higher achievement.

Several contributing factors might explain why critiques received such high response rates despite the longer wait time for responses. When responding to critiques, students needed more time to reflect and possibly conduct research to acquire sufficient information to compose an appropriate rebuttal. At the same time, students probably felt the necessity to post rebuttals to opposing criticisms given the rules and expectations of

the debate activity. In addition, the posting of rebuttals may have been facilitated by the use of message labels, which helped to bring challenges from the opposing team to students' attention. This in turn enabled students to respond to challenges that could be posted anywhere among the many discussion threads.

Graphing the discussion threads across time revealed instances where threads remained active despite the presence of one or more substantially long wait times. For example, message #5 initiated a thread that consisted of 12 responses. In this thread there was a 2.5-day wait time before a response was posted to the opening message. There was another 2.5-day wait time for a response to the third message. But after the fifth day, there was a burst of activity in the discussion thread. A close examination of the text indicated that the first three messages were able to sustain the thread despite the long wait times because the debate teams were exchanging counterarguments (ARGo —> ARGs —> ARGo). This series of challenges set the stage for further discussion and illustrates the significant effect of message content, particularly content that generates conflict and disagreement, on the growth of discussion threads.

The other findings in Table 2 reveal that the overall average wait time across all message categories was .82 days (STD = 1.04). For each response category the average wait time for responses to arguments was .99 days, evidence was .96 days, critiques was 1.04 days, elaborations was .68 days, evaluations was .74 days, and other comments was .50 days. The longest wait time was observed in responses to critiques (1.04 days), which was significantly longer than the wait time for responses to elaborations (T-test=0.038, alpha=.05) and other comments (T-test=.005, alpha=.05).

One explanation for the significant differences in wait times between critiques versus elaborative and other comments is that elaborative and other process comments required less preparation and cognitive effort than posting responses to arguments and critiques. For example, elaborative messages were intended mainly to clarify previous statements, which did not necessarily require the search for additional information or data to ground the response. At the same time, the wait time for responses to critiques was not significantly longer than the wait time for responses to arguments, evidence, and evaluation. This finding suggests that responding to these particular types of messages required approximately the same level of cognitive effort to formulate and post an adequate response.

Effects of response time on response rates. Table 2 shows the mean response times for each type of message: the mean time students needed to compose and post a given type of message (lag 1 event) in response to a previous message (lag 0 event). The mean amount of time taken to respond with arguments was .94 days, .98 days to respond with evidence, .96 days to respond to critiques, .96 days to respond with elaboration, .74 days to

respond with evaluations, and .67 days to respond with other comments. The overall mean time to respond to any given message was .88 days. No significant differences were found between these response times. However, the findings suggest that evaluative and other comments may require the least amount of time to post in response to other messages.

To determine if response times had any effect on the response rates of each message type, the response rates were computed for each type of message across fixed time intervals (see Table 4). For example, arguments (n=40) posted in response to a previous message within the first 12 hours (between 0.0 and 0.5 days) produced a response rate of .68. Arguments (n=18) that were posted in response to a previous message between the first 12 to 24 hours (between 0.5 and 1.0 days) produced a response rate of .50. Overall, the response rates of arguments fell rapidly as the amount of time it took to reply with arguments increased, starting from .68 and dropping steadily to .50, .29, and .43. In contrast, critiques received a response rate of .84 (vs. .50 for arguments) when posted as a reply within 0.5 days. The response rates appear to decline with respect to response time at the same rate as arguments, with response times dropping from .84 to .71, .56, and .56. Nevertheless, critiques maintained a higher overall response rate than arguments.

To compare the rate of decline in response rates over response time between the message categories, the results in Table 4 were plotted onto a graph. The response rates of messages posted with response times of two days or more were not included in the graph due to an insufficient number of messages posted at later time intervals. An analysis of the graph showed that the rate of decline in response rates did not vary significantly across different message content. In other words, the effect of message content did not change the effect of response time on the response rates of

Table 4
Response Rates Across Response Time Intervals for Each Message Category

Days	ARG	EVID	CRIT	ELAB	EVAL	OTH	Mean
0.0	.68	.66	.84	.74	.59	.47	.65
0.5	.50	.50	.71	.75	.59	.46	.60
1.0	.29	.40	.50	.57	.50	.00	.38
1.5	.43	.00	.56	.71	.40	.33	.44
r	−.76	−.95	−.89	−.39	−.94	−.51	−.85
slopw	−.19	−.42	−.21	−.05	−.13	−.18	−.17

r = Pearson correlation between response time (in days) and response rates.

messages. For all message categories combined, the response rates dropped an average of .17 for each day of wait time. No significant differences were found in the test for differences in the response rate by response time slopes for critiques against the slopes of the other message types. However, more data points are needed to produce a more accurate test for potential differences in slopes or rate of decline in response rates. Although the rates of decline in response rates over time were not found to be significantly different between message categories, the graph suggests that some types of messages such as critiques and elaborations may take two or more days to a response rate of zero versus 1.0 to 1.5 days for other message categories. These findings provide further indications of how critiques and elaborative responses can help maintain active discussion threads.

Effects of message-response exchanges on length of discussion threads. Specific message-response exchanges were found to generate higher response rates and more subsequent replies than other types of exchanges. For example, ARG—>CRIT exchanges (*n*=40) elicited a total of 115 subsequent replies within the discussion thread. As a result, each ARG—>CRIT exchange elicited on average 2.88 subsequent replies that contributed to the growth of a discussion thread. Note that CRIT messages when examined in isolation to preceding messages in Table 2 were found to produce a lower response rate of .72. When critical messages were posted in response and in context to arguments, the response rates increased to .80.

In contrast, the response rate for ELAB—>ELAB exchanges were lower at .67, and produced on average 2.70 subsequent replies within discussion threads. The lowest response rates and the lowest number of subsequent replies were found in ELAB—>OTH exchanges with response rate of only .23 and producing only an average of .23 replies. In addition, different levels of conflict can be found in examining the interactions where students respond back to critiques (CRIT—>Reply), possibly to resolve conflicts, respond to disagreements, and/or to bring a thread to its natural conclusion. For example, CRIT—>CRIT interactions produced a response rate of .70 and contributed on average 3.65 subsequent replies to its discussion threads, substantially higher than the CRIT—>OTH exchange with response rate of only .36 and 1.18 subsequent replies.

One explanation for the differences in response rates between the various types of interactions is that ARG—>CRIT and CRIT—>CRIT interactions produced higher levels of conflict or disagreement than ELAB—>CRIT and CRIT—>ELAB interactions. Under the assumptions of dialogic theory, conflict and disagreements in social exchanges generate more critical discourse. The difference in response rates and number of subsequent replies between these interactions is consistent with the assumptions of dialogic theory. Because CRIT messages were found to generate high response rates

despite longer wait times, there is good reason to believe that there are *interactions* like ARG—>CRIT that can also generate higher response rates across longer wait times. As a result, these particular types of interactions should be encouraged in order to mitigate the negative effects of long response times and to sustain the growth of discussion threads.

Implications

Instructional implications. The findings in this study support two strategies for eliciting the types of interactions that minimize the negative effects of long response times that occur frequently in asynchronous discussions. The relatively high response rates observed in this study support the use of debate structures (with assigned team positions) to generate active discussions. The debate format helped encourage and legitimize argumentative exchanges that can withstand long wait times in responses without reducing response rates that can prematurely end discussion threads. The effects of assigning students to teams to defend an assigned position compared with the effects of not assigning students to teams and positions will need further investigation. Nevertheless, the findings in this study strongly suggest the incorporation of argumentative interactions into other types of group activities such as group problem-solving (Cho & Jonassen, 2002), collaborative writing, and peer-evaluation in order to generate active and potentially constructive discussions in online learning environments.

Another factor that may have facilitated argumentative exchanges and sustained discussion is the use of message labels. In this study, message labels are believed to have helped make argumentative exchanges explicit and helped to bring them to students' attention without the assistance of an instructor or moderator. The ability to quickly scan and search for critiques (e.g., CRITo or CRITs) posted by the opposition and locate multiple counter-exchanges between opposing teams (e.g., ARGo—> CRITs and CRITo—> EVIDs) may have helped bring unresolved threads to the attention of students. The use of message labels also enabled students to use the Blackboard search function to query messages by label to monitor the content of the discussions (e.g., count number of arguments posted by the opposing vs. supporting team). The extent to which message labels and constraint-based argumentation can facilitate active threaded discussions needs to be examined under controlled experiments (Jeong, 2000b).

Limitations of the study. Further research is needed to replicate the findings reported in this study because they were based on a limited number of messages generated in a small sample population. Specifically, larger datasets are needed to produce sufficient cell frequencies for the event sequence analysis. Larger datasets will enable a more accurate test of potential differences in the effects of response time on varying message

content across a larger range of response time intervals, as revealed in Table 4. Furthermore, more data will enable statistical tests for differences in the effect of varying student interactions (e.g., ARG—>CRIT vs. CRIT—> CRIT), and not just between specific types of messages examined in isolation. In addition to acquiring larger datasets, the accuracy of students' message labels will need to be improved by introducing prior instruction and training on the labeling process. Finally, the effects of message labeling—a technique used in this study to establish the message as the unit of analysis and to code potentially larger datasets—and the constraints it places on group interaction and response rates will need to be evaluated before continuing its use in future research on group interaction.

Implications for future research. Despite some of the limitations, the software tools described in this study combined with the techniques of student-labeled messages addresses some of the major methodological challenges in CMC research (Rourke et al., 2001). These challenges include difficulties in establishing the unit of analysis in computer conferencing messages, achieving sufficient inter-rater reliability in message coding, finding the necessary time and resources to code sufficiently larger number of conferencing messages, and the inability to map message sequences due to the problems of establishing units of analysis within threaded messages and responses. The tools and techniques developed in this study provide a practical means of analyzing message-response sequences (and the temporal relationships between messages and responses) and analyzing larger datasets needed to conduct rigorous and controlled experiments with multiple treatment groups.

The use of the methods outlined in this study present a number of possible directions for future research. The research on factors that affect student interactions in online environments (including instructor-student, female-male, student-moderator interactions) can be grouped into six categories: time constraints, response constraints, communication styles, student constraints, instructional constraints, and technology constraints. Note that this study examined the combined effects of time constraints and response constraints. Future studies can examine these factors in many possible combinations with or without respect to time constraints. Some examples of the possible research questions are the following: What are the effects of response constraints such as the effects of message content on response rates when message labels are used versus when labels are not used? What are the effects of communication styles such as the use of qualifiers versus intensifiers in arguments and criticisms? What are the effects of student constraints such as differences between men and women in their tendency to post or respond to critiques or the amount of time needed to compose responses? What are the effects of instructional constraints such as the nature of the debate question that allows for disagree-

ment and sharing of multiple viewpoints or the type of group activity such as problem-solving and peer evaluation? What are the effects of technology constraints such as displaying threads in chronological order versus by length of threads or using synchronous versus asynchronous communication tools?

From a broader perspective, this study illustrates the application of event sequence analysis (or interaction analysis) as a method for advancing CMC research on group interaction. Content analysis, as opposed to interaction analysis, is a method that is used increasingly in CMC research (Fahy, Crawford, & Ally, 2001). However, the main limitation of content analysis is its componential approach to analyzing messages in isolation and not in relationship to other messages. Simply counting message/response frequencies across message categories fails to describe the functional and temporal relationships between exchanged messages and responses. Examining and measuring these relationships is needed to better understand the purpose as well as the consequences of engaging students in interactive discourse. New and more sophisticated methods and theories are needed to advance the research in CMC (Koschmann, 1999; Fahy et al., 2001). The methods outlined in this study provide a new framework that will enable researchers to define, measure, and study complex interactions and group processes operationally in online environments (Jeong, 2003a). The ability to measure processes will ultimately enable CMC researchers to conduct more rigorous empirical research to identify the key factors that affect group interaction, the patterns of interaction that lead to improved learning outcomes, and the interventions that produce the desired types of interaction that are proven to support desired outcomes.

References

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ.* Cambridge, UK: Cambridge University Press.

Bakeman, R., & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis.* Cambridge, UK: Cambridge University Press.

Bakhtin, M. (1981). *The dialogic imagination* (M. Holquist, Ed.) Austin, TX: University of Texas Press.

Berge, Z. (1997). Characteristics of online teaching in post-secondary, formal education. *Educational Technology, 37*(3), 35-47.

Cerbin, B. (1988). *The nature and development of informal reasoning skills in collect students.* (ERIC Document Reproduction Service No. ED 298 805)

Cho, K., & Jonassen, D. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3), 5-22.

Duffy, T.M., Dueber, B., & Hawley, C.L. (1998). Critical thinking in a distributed environment: A pedagogical base for the design of conferencing systems. In C.J. Bonk & K.S. King (Eds.), *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse* (pp. 51-78). Mahwah, NJ: Erlbaum.

Fahy, P.J., Crawford, G., & Ally, M. (2001). Patterns of interaction in a computer conference transcript. *International Review of Research in Open and Distance Learning: 2*(1). Available: http://www.icaap.org/iuicode?149.2.1.4

Garrison, D. (1992). Critical thinking and self-directed learning in adult education: An analysis of responsibility and control issues. *Adult Education Quarterly, 42*(3), 136-148.

Gunawardena, C., Lowe, C., & Anderson, T. (1997). Analysis of global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*, 397-431.

Harasim, L. (1993). Collaborating in cyberspace: Using computer conferences as a group learning environment. *Interactive Learning Environments, 3*(2), 119-130.

Hewitt, J. (2003). *Toward an understanding of how threads die in asynchronous computer conferences.* Paper presented at the American Educational Research Association conference, Chicago. Retrieved from: http://www.tigersystem.net/aera2003/revviewproposaltext.asp?propid=6179

Hewitt, J., & Teplovs, C. (1999). An analysis of growth patterns in computer conferencing threads. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of the computer support for collaborative learning (CSCL) 1999 Conference, Dec. 12-15, Stanford University, Palo Alto, California.* Mahwah, NJ: Erlbaum. Retrieved from http://kn.cilt.org/cscl99/A29/A29.htm

Jeong, A. (2003a). The sequential analysis of group interaction and critical thinking in online threaded discussions. *American Journal of Distance Education, 17*(1), 25-43.

Jeong, A. (2003b). *The effects of response constraints and message labels on group interaction and argumentation in online threaded discussions.* Paper presented at the 2004 American Educational Research Association conference.

Koschmann, T. (1999). *Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration.* Paper presented at the computer support for collaborative learning (CSCL) conference.

Levin, J., Kim, H., & Riel, H. (1990). Analyzing instructional interactions on electronic message networks. In L. Harasim (Ed.), *Online education* (pp. 185-213). New York: Praeger.

McAlister, S. (2003). *Assessing good educational argumentation.* Paper submitted to Doctoral Consortium CSCL, 2003, January. Retrieved August 2003 from http://iet.open.ac.uk/pp/s.r.mcalister

Newman, D., Johnson, C., Cochrane, C., & Webb, B. (1996). An experiment in group learning technology: Evaluating critical thinking in face-to-face and computer-supported seminars. *Interpersonal Computing and Technology, 4*(1), 57-74.

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.

Sloffer, S., Dueber, B., & Duffy, M.T. (1999, January). *Using asynchronous conferencing to promote critical thinking: Two implementations in higher education.* Paper presented at the 32nd conference on System Sciences, Maui.

Suthers, D. (1998). *Representations for scaffolding collaborative inquiry on ill-structured problems.* Paper presented at the 1998 AERA annual meeting, San Diego.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*(1), 69-95.

Webb, B., Newman, D., & Cochrane, C. (1994). Towards a methodology for evaluating the quality of student learning in a computer-mediated-conferencing environment. In G. Gibbs (Ed.), *Improving student learning: Theory and practice.* Oxford, UK: Oxford Centre for Staff Development, Oxford Brookes University. 1st International Symposium Improving Student Learning: Theory and Practice, Warwick University, September 1993.

Allan Jeong is an assistant professor of instructional systems in the Department of Educational Psychology and Learning Services at Florida State University. He teaches online courses in a master's program in distance learning with interests in examining and developing methods to examine group interaction and process models in computer-supported collaborative learning environment. He can be reached at jeong@coe.fsu.edu.