

Scaffolding collaborative argumentation in asynchronous discussions with message constraints and message labels

Allan Jeong ^{a,*}, Sunyoung Joung ^{b,1}

^a *Instructional Systems Program, Florida State University, 305E Stone Building, Tallahassee, FL 32306-4453, USA*

^b *Richard Stockton College of New Jersey, P.O. 195 Jimmy Leeds Rd., Pomona, NJ 08240, USA*

Received 18 August 2004; accepted 10 February 2005

Abstract

This study examined the effects of message constraints and labels on collaborative argumentation in asynchronous online discussions. Thirty-eight undergraduate students in an introductory educational technology course were assigned to one of three groups. In one group, students posted specific types of messages using a prescribed set of message categories such as argument, evidence, critique, and explanation. Using the same message categories, another group inserted message labels directly into the subject headings to identify each message by category and increase the visibility of the arguments and challenges presented in debates. A control group received none of the above instructions and constraints. Students in the constraints-with-labels group were significantly less likely to (a) challenge other students, and (b) respond to challenges from other students. The label used to identify critiques might have discouraged students from posting critiques and shifted attention to posting more arguments, following up explanations with more explanations, and evidence with more evidence.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Computer-mediated communication; Distance education and telelearning; Evaluation methodologies; Cooperative/collaborative learning; Pedagogical issues

* Corresponding author. Tel.: +1 850 644 8784; fax: +1 850 644 8776.

E-mail addresses: jeong@coe.fsu.edu (A. Jeong), joungs@stockton.edu (S. Joung).

¹ Tel.: +1 609 652 4559.

1. Introduction

Computer-mediated communication (CMC) has been shown to increase learner-to-learner interaction and facilitate critical thinking in online group discussions (Collins & Collins, 1996; Ravits, 1997; Ward & Tiessen, 1997). As a result, CMC has been used as a means to support collaborative argumentation (CA) – a collaborative inquiry-based learning strategy to practice and develop critical thinking skills in online settings (Derry, Levin, & Osana, 2000). The skills of argumentation involve the processes of building arguments to support a position, and considering and weighing evidence and counter-evidence in developing supporting arguments and identifying weaknesses in opposing claims. Argumentation is also a process for testing out uncertainties, to extract meaning and to achieve deeper understanding (McAlister, 2001). In collaborative argumentation, students discuss and examine complex ill-structured problems and to acquire the skills of argumentation and reasoning, students must practice the skills through discussion and in written scholarship (Lipman, 1991). Collaborative argumentation in CMC or computer-supported collaborative argumentation (CSCA) provides students the opportunity to practice argumentation through writing and discussion simultaneously as students communicate online using text-based communication tools (Baker, 1999).

One of the approaches to scaffolding collaborative argumentation examined in CSCA research is to embed constraints within the discussion environment and interface which requires students to classify and label each posted message to a specific functional category (e.g., argument, evidence, critique, explanation). In this approach, students' options on the types of messages and responses they can post to a discussion are restricted to a prescribed and stringent set of message categories presented within the discussion environment. The primary purpose for placing constraints on students' messages is to assist students in maintaining a task-oriented discussion based on some presumed model of the argumentation process. Its second purpose is to reduce the difficulty of teasing out the structure of arguments raised in discussions and to support better understanding of task-related mental structures and arguments (Derry et al., 2000). The ability to better visualize the complex discourse structures can ultimately support problem solving, group communications and learning (Buckingham-Shum, MacLean, Bellotti, & Jammond, 1997; Cho & Jonassen, 2002; Suthers, 1998; Suthers & Hundhausen, 2001; Suthers, Hundhausen, & Girardeau, 2003; Tan, 2000; Turoff & Hiltz, 1998; Veerman, Andriessen, & Kansellaar, 1999). Finally, message constraints and labels can be used to facilitate analysis of the sequential nature of exchanged messages and enable a process-oriented approach to studying student interaction in online discussions (Jeong, 2003a, 2003b, 2003c, 2004).

Despite these claims, various criticisms and concerns have been or should be raised about the use of message constraints and message labels, particularly when the constraints are embedded within a CMC environment. Some of these concerns are that: (1) too much structure reduces level of dialogue and student interaction (Saba & Shearer, 1994), and the constraints inhibit group spontaneity, brainstorming and idea generation given that CMC by itself, without message constraints, already inhibits these particular processes (Hiltz, Turoff, & Johnson, 1986; Hollingshead, 1996); (2) too many restrictions are placed on what students can and cannot contribute to discussions, and that such approaches can appear to students to be overly coercive (Beers, Boshuizen, & Kirschner, 2004); (3) ideas cannot always be assigned explicitly to a given label, and as a result, some ideas may be suppressed or omitted from discussion; (4) ideas cannot be adequately discussed by separating them into separate discrete messages because students already experience

difficulty in seeing the relationships between ideas presented in discussions (Sproull & Keisler, 1991); (5) the formalized procedures represented and mediated through the message constraints might be incomplete, inadequate or even entirely inappropriate based on the argument that actions and tasks are situated and determined by context (Schmidt, 1999); and (6) the computerization of processes can make the procedures problematic because groups must adjust to the computerized methods before the group can function and focus on the task (Poole, Holmes, Watson, & DeSanctis, 1993).

The following is a description of the message constraint and message label approach to scaffolding collaborative argumentation and a discussion of findings from previous studies that reveal its instructional benefits and challenges. Also discussed are previous studies that have used message constraints and labels as a means to measuring how various factors affect the processes of collaborative argumentation in threaded discussions. The findings highlighted in this discussion provide the rationale for the specific research questions addressed in this study.

1.1. Instructional applications for message constraints and labels

Numerous text-based communication tools have been developed to support collaborative argumentation by embedding message constraints and rules of argumentation within the discussion environment (Duffy, Dueber, & Hawley, 1998; Jonassen & Remidez, 2002; Leinonen, Virtanen, & Hakkarainen, 2002; McAlister, 2003; Weinberger, Fischer, & Mandl, 2001). For example, Jonassen and Remidez (2002) developed a threaded discussion tool called ShadowPDforum that maps specific argument structures onto the hierarchically organized messages by requiring students to classify the function of their messages (e.g., claim, supporting evidence, rebuttal) before messages are posted to the discussion. In this particular system, constraints are also placed on message–response sequences such that messages are attached to responses by a set of constrained links so that, for example, claims can only be linked to supporting evidence, and counter claims can only be linked to rebuttals. The technique of placing constraints on what types of messages can be posted to a discussion, and the use of labels to mark the function of each message, has been applied in other asynchronous discussion environments like ACT (Duffy et al., 1998; Sloffer, Dueber, & Duffy, 1999), FLE3 (Leinonen et al., 2002), Ntool (Beers et al., 2004), and in synchronous chat tools like AcademicTalk (McAlister, 2003).

Weinberger et al. (2001) used message constraints and labels to implement “cooperative scripts” designed to assist students in performing the roles of analyzer and constructive critique. Multiple message categories were presented to help students produce messages that served a particular role (e.g., various types of critiques). They found that cooperative scripts produced more equal participation among students within discussion groups, greater gain in individual students’ knowledge transfer, and higher convergence in shared inferences between group members. Weinberger, Fischer, and Mandl (2004) also found that cooperative scripts significantly increased the divergence of presented knowledge and viewpoints, but had no effect on the groups’ ability to converge towards shared consensus.

Stegmann, Weinberger, Fischer, and Mandl (2004) used a synchronous chat environment to study the effects of message constraints and labels (e.g., claim, premise or restrictions and warrants) and the effects of constraints on response sequences when messages were automatically pre-set and labeled as argument, counter argument, or integration depending on the type of

message eliciting the reply. The findings showed that students generated fewer unsupported claims and achieved greater knowledge of the argumentation process. No differences were found in individual knowledge acquisition, students' ability to apply relevant information and specific domain content to arguments, and ability to converge towards shared consensus.

1.2. Research applications for message constraints and labels

A few recent studies used message constraints and labels as a means to operationalize and measure the frequency of specific message–response sequences to examine the functional, temporal and social relationships between messages and responses in asynchronous threaded discussions (Jeong, 2003b, 2003c, 2004). In these studies, message constraints and labels were used specifically to address one of the major challenges in coding and analyzing online discussions (Garrison, 2000; Rourke, Anderson, Garrison, & Archer, 2001). The challenge is that online messages often address multiple functions at the same time, particularly in asynchronous discussions when more time is available to compose messages of greater length, making it very difficult to establish the unit of analysis (Gunawardena, Lowe, & Anderson, 1997; Levin, Kim, & Riel, 1990; Newman, 1995). Restricting each message to serve only one function at a time enabled these studies to examine group interaction by using message–response pairs as the unit of analysis. As a result, the sequencing of messages and responses were successfully mapped by determining precisely how likely one type of message was able to trigger a particular type of response. For example, the studies measured the probabilities of eliciting critiques when posting arguments, and the probabilities of eliciting explanations when posting critiques to arguments.

In one case study, Jeong (2004) examined how the combined effects of message function and response time determined how likely a response elicited a reciprocal response from other students. The findings showed that in general the longer a student waited to post a reply to another student's message, the less likely the student's reply was to elicit a reciprocal reply from other students. However, the main finding of this study was that critical responses to presented arguments elicited the highest response rate among the other possible types of responses regardless of how long a student took to post the critique in response to a previously stated argument. These findings suggest that the increased visibility of critical responses resulting from the use of message labels can help students sustain and advance discussion threads, which are difficult to achieve in asynchronous threaded discussions (Hewitt & Teplovs, 1999).

In another study (Jeong, 2003c), a controlled experimental design was used to test the effects of constraining the sequences of messages and responses to conform to a particular model of argumentation. For example, arguments could only be followed by replies presenting critiques or evidence, and could not be followed by arguments supporting the opposing position. In other words, the sequential constraints prohibited students from simply replying to main arguments with opposing main arguments or replying in ways that did not directly address issues specifically raised in a previous message linked to the reply. By reducing the types of replies that were not conducive to advancing a discussion thread or argument, the effects of the sequential constraints were found to significantly increase the number of replies that were deemed to be productive. Specifically, this study found a significant increase in the number of elaborations and explanations posted in reply to critiques. These results helped students advance their discussions, and enabled them to exhibit higher-order thinking and reasoning.

In a third study, the gender of a participant was found to be useful in predicting how and how likely students responded to messages from males versus females (Jeong, 2003b). Females were found to be more likely to critique a message from males than from females, males were more likely than females to engage in extended exchanges of critiques, rebuttals and elaborations, and males had a strong tendency to respond to female arguments with supporting evidence. These findings suggested that group composition can strongly influence argumentation between students and that different ways of presenting message constraints and ways to label messages may be needed to encourage specific behaviors not often exhibited by participants of a certain gender.

1.3. Limitations of previous research

One major limitation of the previous studies on CSCA is that little information is provided on how to refine prescribed message categories and message labels to further facilitate argumentation because these studies were primarily descriptive than prescriptive in nature. Previous studies were descriptive because they only examined the relative frequencies of students' postings for each message type. A prescriptive approach can be achieved by measuring the response probabilities and response distributions for each message type, thus providing a means to predict how likely arguments elicit responses versus no responses, and when they do elicit responses, how likely is the response to be a critique, supporting evidence, explanation, or counter-argument. This type of analysis provides the means to develop process models of collaborative argumentation that examine message–response sequences or patterns of interaction that exemplify productive argumentation (e.g., “argument → critique → explanation” and “argument → critique → counter-critique”). Surprisingly, previous studies report little or no data on how likely arguments elicit supporting evidence or critiques, and how often critiques elicit explanations and counter-arguments. New studies are needed to systematically identify message constraints and effective ways to label the messages to elicit the types of responses that support rather than inhibit argumentation.

Although some preliminary studies have examined the processes of argumentation using message constraints with labels and a prescriptive approach (Jeong, 2003b, 2003c, 2004), the findings from these studies cannot yet be generalized because the findings may have been the direct result of using message constraints and message labels. For example, the effects of gender in collaborative argumentation (Jeong, 2003b) were not tested under conditions using message constraints with message labels versus message constraints without labels. What is uncertain at this time is whether or not the significantly lower rate of confrontational exchanges between females were the product of having to explicitly label messages using the label “critique”, and as a result, drawing greater and unwanted attention to the conflict produced by their critiques. Would the males and females have produced the same response patterns if message constraints were used without the “critique” label or used with a label that carried less or no negative connotation (e.g., “but” or “question”)?

These noted limitations of the previous studies suggest that a controlled study is needed to test for differences in the effects of using message constraints with labels versus without labels in order to (a) understand their relative impact of how message constraints and message labels affect student interaction patterns, and (b) how these resulting patterns support the processes that lead to higher levels of critical discourse. The main rationale for conducting this comparative study is to test the possibility that how and when students respond to messages may be largely affected by

how messages are labeled and not by how messages are constrained to specific functions. This controlled comparison would then lay the foundation to conduct further studies on the effects of gender, response time, or any host of factors that are believed to affect the processes of argumentation in online discussions.

1.4. Theoretical assumptions and rationale for the study

The rationale for comparing the effects of message constraints with labels versus without labels on message–response sequences and for using a process-oriented approach for comparing their effects was based on the assumptions of Dialogic theory of language (Bakhtin, 1981; Koschmann, 1999; Sappo & Mononen-Aaltonen, 1998). The theory assumes that social dialogue must be viewed as part of a 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. The second assumption is that meaning is renegotiated and reconstructed as a direct result of *conflict* in social interactions. This assumption is supported by extensive research on collaborative learning that has shown that conflict and the consideration of both sides of an issue is needed to drive inquiry, reflection, articulation of individual viewpoints and underlying assumptions, and to achieve deeper understanding (Johnson & Johnson, 1992; Wiley & Voss, 1999).

The implications of these assumptions are that conflict (e.g., argument → critique) and subsequent inquiry triggered by conflict (e.g., critique → explanation, critique → evidence) is produced not by the ideas presented in one message alone, but by the inter-relationship between ideas presented in a message and responses to the message. Given these assumptions, there is reason to believe that message constraints and message labels can significantly affect how often particular types of interactions occur in online discussions. For example, explanations posted in rebuttal to critiques could increase in frequency when “explanation” is explicitly presented as a response option. Furthermore, explanations could also increase in frequency when message labels are used to increase the saliency of critiques and as a result increase the ability of students to identify and respond to arising conflicts.

1.5. Purpose of study

The purpose of this study was to determine how message constraints and message labels affect how and how often students respond to specific types of messages to support argumentation. This study addressed two basic questions regarding two types of message–response sequences needed to produce and higher levels of critical analysis:

- (1) *Challenging other students* – Do message constraints with labels increase or decrease the mean number of times a student critiques and challenges messages posted by other students (e.g., argument → critique, evidence → critique)?
- (2) *Responding back to challenges* – Do message constraints with labels increase or decrease the mean number of times a student responds back to critiques with rebuttals to provide explanations, counter-critiques, and additional evidence to defend previous claims (e.g., critique → explanation, critique → counter-critique)?

2. Methods

2.1. Participants

In this study, 38 pre-service teachers participated in a series of five debates on five educational issues that took place once per week over a period of 5 weeks. The students were largely undergraduates enrolled in a course on educational technology with 6 males and 32 females.

2.2. Treatment groups

Students were randomly assigned to three treatment groups – control group, constraints-only group, and constraints-with-labels group. In the control group, students received no instructions on what types of messages to post to the debates. Due to attrition, 10 students participated in the control group with 2 males and 8 females.

In the constraints-only group, students were instructed to (a) post specific types of messages from a prescribed set of message categories such as arguments, evidence, critique, and explanation (see Fig. 1), and (b) restrict the content of each message so that each message addressed only one category at a time. The message categories were based on Toulmin's (1958) model of argumentation and derived from a previous study that conducted a content analysis of on-line debates in a MBA course (Jeong, 2003a). This group consisted on 12 females and 2 males.

ARG	Argument in support of the proposition. For example, "Computers can substitute the role of the instructor when kids use computer-based tutorials". The subject heading for this argument might be something like "ARGs computers as tutors"
EVID	Evidence, specific examples, research studies, personal observations & experiences, or proofs. For example "I found an article that shows that..." or "In one course, I used a computer tutorial and I found that...". The subject heading for these messages might be "EVIDs study finds tutorials to be effective".
CRIT	Challenge or critique given arguments, evidence or any other responses from the opposing team. For example, "The personal experiences you've had with computer-based tutorials cannot be generalized to other students because" or "The study you cited was flawed because..."
EXPL	Explanation, elaborations or clarifications. For example "Computer-based tutorials perform the specific role of assessing students understanding just as an instructor can administer pop quizzes in a classroom".
OTH	Other process comments or questions or requests for help, such as "What is a computer-based tutorial?" or "Thanks for responding to my message" or "I think we need to provide more evidence, such as journal articles, textbook readings, personal observation & experiences from past courses, to show that computer-based tutorials perform just as well as an instructor?" or "What exactly did you mean when you said that...?" or "Can you give me more evidence?"

Students were also required to insert tags ('o' = opposing, 's' = supporting) at the end of each label to identify postings from the opposing versus the supporting team. For example, ARGs for an argument from the supporting team and CRITo for a critical response from the opposing team.

Fig. 1. Message categories used to scaffold and label messages during argumentation.

Students in the constraints-with-labels group, consisting of 12 females and 2 males, were instructed to (a) post specific types of messages using the same prescribed set of message categories used in the constraints-only group, (b) restrict the content of each message so that each message addressed only one category at a time, and (c) label each message with a prescribed label assigned to each message category. Students were required to manually insert the labels into the subject headings of each message (see Fig. 2). Although message titles were not required to be included with the labels, 68% of messages in the constraints-with-labels group were posted with message titles, compared to 32% and 36% in the control and constraints-only group, respectively. Regardless, students in the constraints-with-labels group were able to see all message labels on screen while reading the discussion threads. Therefore, the students were able to see the hierarchical structure of their arguments presented in the hierarchically structured discussion threads. The experimenter occasionally checked the message labels for errors in students' labels and instructed students to return to the messages to correct for errors. No participation points were awarded for a given debate when students failed to properly label a message.

This study did not compare the effects of message constraints without labels versus message labels without message constraints because a prescribed set of message labels is equivalent to

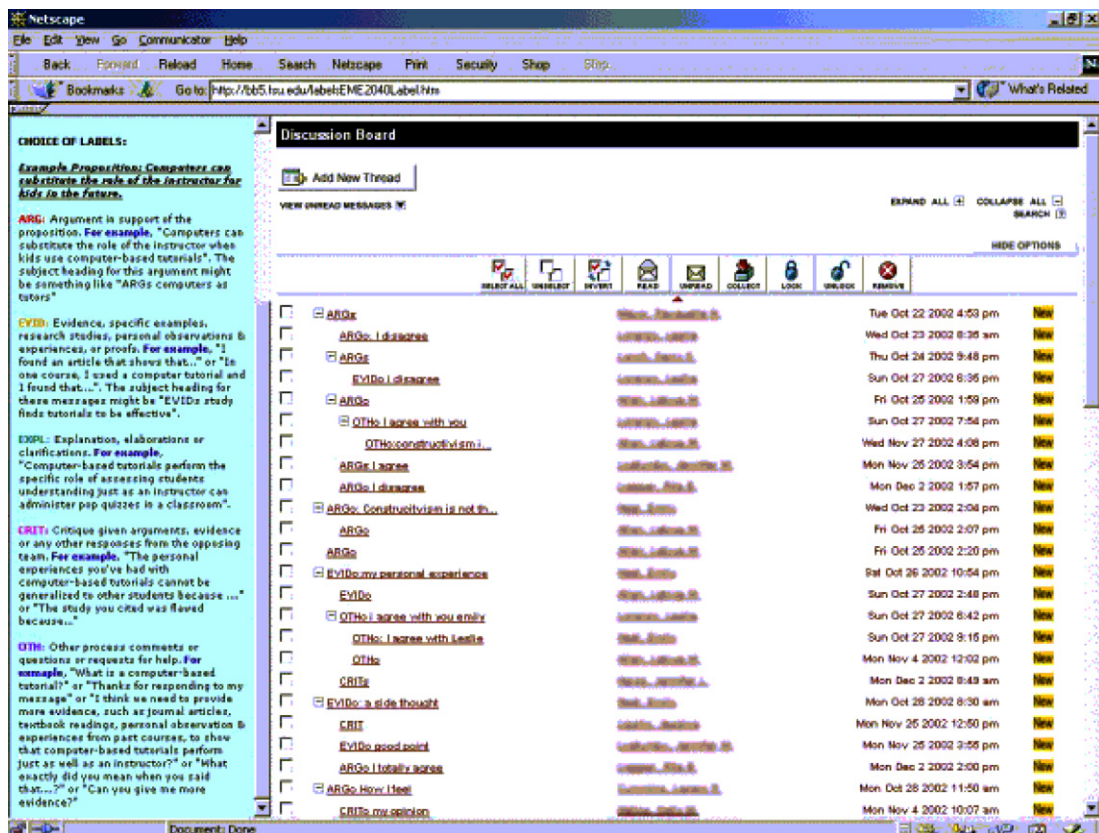


Fig. 2. Illustration of a threaded discussion forum from the constraints-with-labels group with prescribed message categories and labels displayed in the left panel.

imposing constraints on messages given that each label is implicitly associated with a specific function or message category.

Students in all the groups were required to post a minimum of four messages per debate in a threaded discussion board in Blackboard, a course management system, to receive four participation points per debate. Within each group, students were assigned to a team to either support or oppose a position. In each debate, students debated issues that explored the appropriate use of educational technology. For example, students debated the position “Computers can substitute the role of the instructor for kids in the future”.

The total number of messages produced in the five weeks of discussion was 207 in the control group, 300 in the constraints-only group, and 265 in the constraints-with-labels group. The mean number of postings per students in the control, constraints only, and constraints-with-labels groups were 4.24 ($SD = 1.16$), 4.29 ($SD = 0.82$) and 4.14 ($SD = 0.49$), respectively. The overall percentage of messages that elicited one or more replies in the control, constraints only, and constraints-with-labels groups were 0.63, 0.60 and 0.56, respectively.

2.3. Statistical analysis

The dependent variable examined in question 1 was the mean number of times a student critiqued (or challenged) messages posted by other students. The dependent variable for examining question 2 was the mean number of times a student responded back to other students' critiques. Each dependent variable was used to test for differences between the control versus constraints-only group, and to test for differences between the constraint-only versus constraints-with-labels group. As a result, a total of four *t*-tests were used to examine the two questions addressed in this study. The Bonferroni adjustment was applied to set the α level at $p = 0.0125$ ($p = 0.05/4 = 0.0125$) for each *t*-test.

This study also used event sequence analysis (Bakeman & Gottman, 1997) to identify prevalent patterns in message–response exchanges. Sometimes referred to as interaction analysis, lag analysis or Markovian chain analysis, the method of event sequence analysis was used to determine: (1) the probability in which a given message was able to elicit a specific type of response; and (2) the probability distribution of the types of responses elicited by each message type. Event sequence analysis has effectively used in other research on group and inter-personal communication to study, for example, communication patterns in the conversations and interactions between married couples (Bakeman & Gottman, 1997, pp. 184–193; Gottman, 1979), children at play (Bakeman & Brownlee, 1982), mother and infant at play (Stern, 1974), and humans and computer-interfaces (Olson, Herbsleb, & Rueter, 1994).

Using event sequence analysis, this study identified prevalent patterns in message–response sequences by: (1) counting the frequency of specific responses to each type of message; (2) converting the response frequencies into transitional probabilities for each observed message–response interaction, e.g., probability of an argument eliciting a critical response versus a response with supporting evidence; and (3) converting the transitional probabilities into transitional state diagrams to provide a visual birds-eye view of interaction patterns. The analysis also included relative frequencies and response rates for each response category. The statistical analysis and transitional state diagrams were generated for each treatment group using the software program, Discussion Analysis Tool (Jeong, 2003a).

Two coders classified the messages in all three groups to check inter-rater reliability and to assign a code to each message posted in the discussions. The Cohen's κ test of the reliability of the codings was very good for the control group ($k=0.82$), the constraints-only group ($k=0.87$), and the constraints-with-labels group ($k=0.87$). A comparison of the students' labels with the labels assigned by one of the coders revealed a reliability of $k=0.31$. The most frequent errors in the message labels occurred when explanations were classified as arguments ($n=10$), explanations as other statements ($n=16$), critiques as arguments ($n=13$), arguments as explanations ($n=10$), and arguments as other statements ($n=9$). Students in the constraints-with-labels correctly labeled 52% of their messages. In all three groups, the experimenter's codes were used for the data analysis.

3. Results

3.1. Challenging other students

Table 1 contains the mean number of times a student critiqued (or challenged) messages posted by other students within each group. Between the control and the constraints-only group, no significant differences were found in the mean number of critical replies, $t(24)=0.22$, $p>0.0125$. Significant differences were found in the mean number of critical replies between the constraints-only and constraints-with-labels groups, $t(27)=3.67$, $p<0.0125$. The mean number of critical replies per student in the constraints-only group was 5.21 per student ($SD=3.31$) compared to 1.80 in the constraints-with-labels group ($SD=1.37$), with an effect size of -0.98 . These findings show that constraints with labels significantly lowered students' tendency to challenge other students' messages.

3.2. Responding back to challenges

Table 2 contains the mean number of times a student responded back to other students' critiques within each group. No significant differences were found in the mean number of replies to critiques between the control versus constraints-only group, $t(24)=-0.62$, $p>0.0125$. Significant differences was found in the mean number of replies to critiques between the constraints-only versus constraints-with-labels group, $t(27)=2.88$, $p<0.0125$. The mean number of replies to critiques in the constraints-only group was 2.43 per student ($SD=1.74$) compared to 0.87 in the constraints-with-labels group ($SD=1.13$), with an effect size of -1.45 . These findings show that constraints with labels significantly lowered students' tendency to respond back to critiques to defend previous claims.

Table 1
Mean number of critical replies posted per student

	Control	Constraints	w/Labels
Mean	4.92	5.21	1.80
SD	3.48	3.31	1.37
<i>N</i>	12	14	15

Table 2

Mean number of replies to critiques posted per student

	Control	Constraints	w/Labels
Mean	1.92	2.43	0.87
SD	2.43	1.74	1.13
N	12	14	15

3.3. Post hoc analysis of message frequencies and distribution

Given the finding that students in the constraints-with-labels group were less likely to critique other students and respond back to critiques, post hoc tests were conducted to determine if this pattern of interaction significantly affected the types of messages students contributed to the debates. For example, Table 3 indicated that 27% of all messages posted by students in the control group were arguments versus 35% in the constraints-only group and 41% in the constraints-with-labels group. To determine if the frequency distributions differed significantly between the groups, three χ^2 tests of independence were conducted in post hoc analysis.

Table 3

Transitional probabilities, number of replies, number of messages per category, response rate, relative frequency of replies elicited per category, and relative frequency of messages per category

	ARG	EVID	CRIT	EXPL	OTH	Reply	Given	Rate	%RPY	%GIV
<i>Control</i>										
ARG	0.06	0.06	0.45	0.42	0.01	108	55	0.82	0.68	0.27
EVID	0.20	0.00	0.20	0.60	0.00	5	10	0.50	0.03	0.05
CRIT	0.04	0.08	0.46	0.42	0.00	24	70	0.60	0.15	0.34
EXPL	0.00	0.05	0.41	0.55	0.00	22	70	0.54	0.14	0.34
OTH	0.00	0.00	0.00	0.00	0.00	0	2	0.50	0.00	0.01
						159	207			
<i>Constraints-only</i>										
ARG	0.07	0.06	0.45	0.41	0.01	122	105	0.70	0.61	0.35
EVID	0.08	0.23	0.23	0.46	0.00	13	16	0.44	0.06	0.05
CRIT	0.06	0.11	0.61	0.22	0.00	36	99	0.55	0.18	0.33
EXPL	0.10	0.00	0.50	0.37	0.03	30	78	0.58	0.15	0.26
OTH	0.00	0.00	0.00	0.00	0.00	0	2	0.50	0.00	0.01
						201	300			
<i>Constraints with labels</i>										
ARG	0.21	0.07	0.23	0.47	0.02	97	109	0.43	0.67	0.41
EVID	0.00	0.36	0.21	0.36	0.07	14	26	0.42	0.10	0.10
CRIT	0.13	0.07	0.40	0.40	0.00	15	37	0.30	0.10	0.14
EXPL	0.22	0.00	0.11	0.67	0.00	18	81	0.17	0.12	0.31
OTH	0.00	0.00	0.00	0.00	1.00 ^a	1	12	0.08	0.01	0.05
						145	265			

^a Note that this transitional probability of 1.00 for OTH → OTH interaction was based solely on a frequency of 1 OTH reply to OTH message.

Significant differences were found in the frequencies of messages across the five message categories between the control versus constraints-with-labels group, $\chi^2(4, N=472)=36.43, p<0.01$. Specifically, the relative frequency of arguments in the constraints-with-labels group (0.41) was higher than the control group (0.27), while the relative frequency of critiques was lower in the constraints-with-labels group (0.14) than the control group (0.34).

Significant differences were also found between the constraints-only versus constraints-with-labels group, $\chi^2(4, N=565)=35.88, p<0.01$. The frequency of arguments in the constraints-with-labels group (0.41) was higher than the constraints-only group (0.35), while the frequency of critiques in the constraints-with-labels group (0.14) was lower than the constraints-only group (0.33). These findings overall show that constraints with labels (as implemented in this study) produced higher frequencies of arguments, possibly because students' made efforts to make up for the lower number of messages posted as critiques and replies to critiques. This finding shows that the constraints with labels used in this study significantly increased the number of proposed arguments considered in the debates, but at the same time, inhibited the process of developing deeper and more critical analysis of individual arguments.

No significant differences were found in the relative frequencies between the control versus the constraints-only group, $\chi^2(4, N=507)=5.54, p>0.01$. This finding suggests that the use of message constraints alone did not contribute to increasing the relative frequencies of posted evidence, critiques and explanations in response to presented arguments.

3.4. Post hoc analysis of specific message–response sequences

To conduct a closer examination of some of the specific processes (or message–response sequences) that can be used to demonstrate or measure higher levels of critical analysis, a post hoc analysis was conducted to test for possible differences between the three groups on the mean number of (a) critical replies to arguments (ARG → CRIT), (b) explanatory replies to critiques (CRIT → EXPL), and (c) critiques or counter-critiques posted in reply to critiques (CRIT → CRIT). The assumption is that the message–response sequences of ARG → CRIT → EXPL and ARG → CRIT → CRIT are two fundamental processes needed to achieve higher levels of critical discussion.

Significant differences were found in the mean number of critical replies to arguments (ARG → CRIT) between the three groups based on a one-way analysis of variance for three independent samples, $F(2,269)=11.95, MSE=0.75, p<0.01$. Table 4 shows that arguments elicited an average of 0.89 critiques in the control group, versus 0.52 critiques in the constraints-only group, and 0.20 critiques in the constraints-with-labels group. The effect size for the differences between the control group and the constraints-with-labels group was -0.76 , and the effect size for the differences between the constraints-only versus constraints-with-labels group was -1.88 . In terms of the proportion of arguments that elicited critiques, 51% of arguments in the control group elicited critical replies compared to 29% in the constraints-only group, and only 16% in the constraints-with-labels group. These findings show that message labeling significantly reduced the frequency of arguments that were critically challenged and analyzed. In other words, constraints with labels significantly affected students' tendencies to critically challenge or not challenge the arguments of other students.

No significant differences were found in the mean number of explanations posted in reply to critiques (CRIT → EXPL), $F(2,206)=0.95, MSE=0.13, p>0.01$. The average number of explanatory

Table 4

Measures of group performance based on mean number of targeted replies to given messages and the percentage of given messages with specific targeted replies

	Control	Constraints	w/Labels
ARG–CRIT			
Mean number of CRIT replies per ARG	0.89	0.52	0.20 ^a
SD	1.17	0.96	0.52
Percent ARG with CRIT replies	0.51	0.29	0.16
CRIT–EXPL			
Mean number of EXPL replies per CRIT	0.14	0.08	0.16
SD	0.39	0.27	0.50
Percent CRIT with EXPL replies	0.13	0.08	0.11
CRIT–CRIT			
Mean number of CRIT replies per CRIT	0.16	0.22	0.16
SD	0.40	0.55	0.44
Percent CRIT with CRIT replies	0.14	0.16	0.13

^a $F(2, 269) = 11.95$, $MSE = 0.75$, $p < 0.0001$.

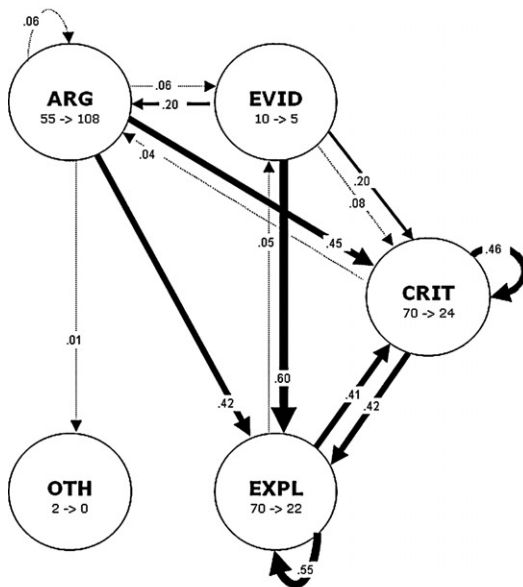
replies to each critique was 0.14 in the control group, 0.08 in the constraints-only group, and 0.16 in the constraints-with-labels group. Only 13% of critiques received any explanatory replies in the control group, 8% in the constraints-only group, and 11% in the constraints-with-labels group.

Furthermore, no significant differences were found in the mean number of counter-critiques posted in reply to critiques (CRIT → CRIT), $F(2, 206) = 0.44$, $MSE = 0.23$, $p > 0.01$. The mean number of critical responses to each critique was 0.16 in the control group, 0.22 in the constraints-only group, and 0.16 in the constraints-with-labels group. Only 14% of critiques received counter-critiques in the control group, 16% in the constraints-only group, and 14% in the constraints-with-labels group. These two latter findings indicate that constraints with labels did not affect students' tendencies to respond or not respond back to critiques to defend a previous claim or argument.

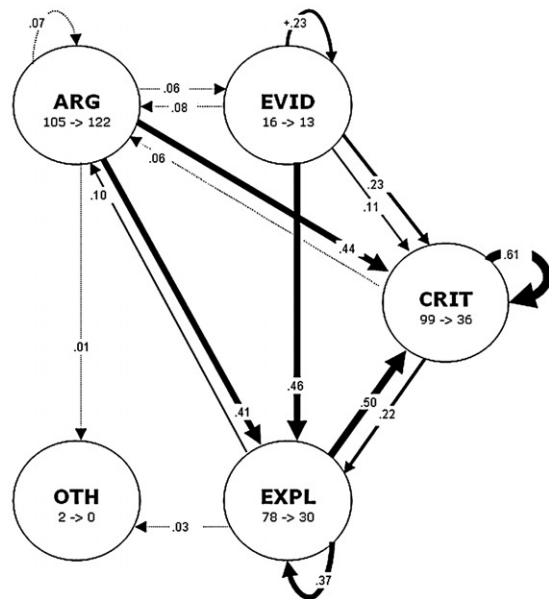
3.5. Exploratory analysis of group response patterns

To identify alternative explanations for these findings, the transitional probabilities in Table 4 were converted into a transitional state diagram for each group (see Fig. 3). The diagrams provide a visual illustration of how and how often students responded to given types of messages when students chose to respond to a message (versus choosing not to respond at all). The patterns in students' response choices suggests that the lower frequency of challenges posted by students in the constraints-with-labels group may be attributed to the students' tendency to respond to arguments with counter-arguments or ARGo → ARGs (0.21 of responses) compared to students in the control group (0.06) and the constraints-only group (0.07). This pattern of responses in the constraints-with-labels group may have diverted students' attention away from responding to arguments with critiques. This particular response pattern, ARGo → ARGs, provides one possible explanation as to why students in the constraints-with-labels group responded with critiques in only 0.23 of their responses to arguments, compared to 0.45 in the control group and 0.45 in the constraints-only group.

Control group



Constraints-only



Constraints-with-labels

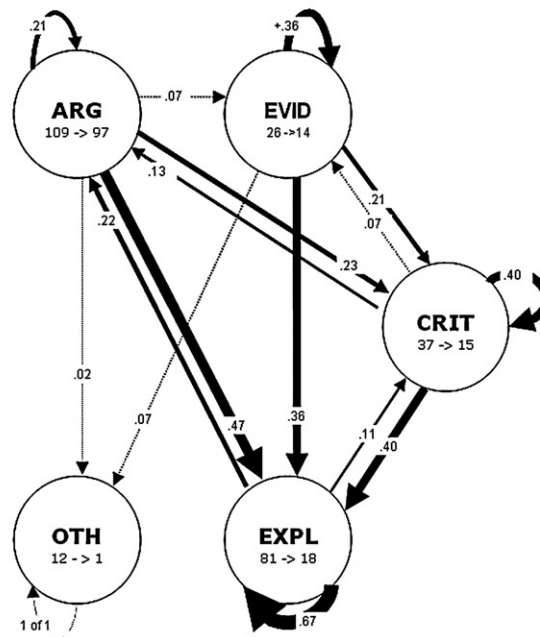


Fig. 3. Transitional state diagrams of the observed message response interactions with transitional probabilities, message frequencies and response frequencies. *Note.* The observed frequencies for each message type and responses to the message are displayed within each node.

Furthermore, the diagrams reveal a higher frequency of $\text{EXPL} \rightarrow \text{EXPL}$ and $\text{EVID} \rightarrow \text{EVID}$ interactions in the constraints-with-labels group, which suggest that students in the constraints-with-labels group focused more of their postings to support rather than challenge one another's postings. These two particular patterns of interaction may also have contributed to the low frequency of challenged arguments observed in the constraints-with-labels group. Finally, observed similarities in the distribution of responses to critiques between all three groups help to explain why no differences were found in the frequency of explanatory replies to critiques and counter-critiques posted in reply to critiques between the three groups.

4. Discussion

The purpose of this study was to determine the efficacy of using message constraints with message labels to produce the types of message–response sequences that demonstrate high levels of argumentation and analysis. The main findings in this study, however, showed that the message labels inhibited the processes needed to produce critical argumentation. This study found that students who used message labels were 2–3 times less likely to challenge other students and 2–3 times less likely to respond back to challenges. According to Baker (1999), the need to explain, justify, or understand is felt and acted upon only when conflicts or errors are brought to attention. Critical responses to messages, particularly to arguments, are essential because this process plays a key role in increasing students' understanding and improving group decision-making (Lemus, Seibold, Flanagan, & Metzger, 2004). As a result, the findings in this study suggest that message labels can potentially inhibit critical argumentation and possibly inhibit student learning.

One plausible explanation for the lower frequency of critiques and $\text{ARG} \rightarrow \text{CRIT}$ exchanges is that the label used in the constraints-with-labels group to identify critiques, "CRIT", carried negative connotations. Its negative connotations could possibly have heightened the perception of critiques and the perceptions of the students that post critiques as being overly confrontational. Butler and Geis (1990) have shown that some behaviors (e.g., critiquing) when exhibited by females can be viewed by others as dominating and emotional, whereas the very same behaviors exhibited by males are often perceived and judged as valuable and responsible contributions to the group. To further compound this problem is that the majority of the students in this study were female, and the research has shown that the female style of communication is more likely to be supportive rather than confrontational (Tannen, 1990), as was demonstrated by the higher frequency of $\text{EXPL} \rightarrow \text{EXPL}$ and $\text{EVID} \rightarrow \text{EVID}$ interactions observed in this study.

In contrast, males are more likely to engage in argumentation, confront differences in opinions, and defend viewpoints (Tisdell, 1993; Vanfossen, 1996). As a result, a higher proportion of males in the debates could have compensated for the effects of using "CRIT" in labeling the critiques. Furthermore, Jeong (2003b) found evidence to show that females are (a) more likely to critique arguments in online discussions when responding to arguments posted by males rather than by females, and (b) just as likely as males to challenge an argument when the argument is posted by a male. As a result, increasing the male-to-female ratio could also help to increase the number of challenges and response to challenges posted by females and not just by males. If this study had been conducted with all-male or mostly male discussion groups, there is the possibility that the findings in this study could have been entirely reversed to show that constraint-with-labels can improve and not inhibit argumentation.

Another plausible explanation for the findings is that the students may have implicitly established a group norm where the students implicitly agreed not to challenge one another's arguments due entirely to isolated events, circumstances or idiosyncratic behaviors of particular members in the group. If this were the case, students in the constraints-with-labels groups would have no other option than to focus more of their efforts on identifying arguments and posting support for the arguments as observed in the post hoc analysis. In this study, students were not required nor explicitly instructed to respond to all arguments with critiques. Furthermore, no constraints were placed on message–response sequences – only constraints on message categories. Also, the students did not receive explicit instructions or models on how to challenge arguments. Therefore, another possible explanation is that the students in the constraints-with-labels group, by chance alone, may have been less skilled in challenging arguments than students in the other groups. All of these explanations, however, will need to be further examined and validated through group interviews and pre-testing students' argumentation skills to control on argumentation skills of students to control for individual differences in argumentation skill.

5. Implications for future research in CMC

Given that the findings of this study are not yet conclusive, future research will need to address some of the following issues. Multiple discussions groups will need to be examined within each experimental condition so that the idiosyncratic behaviors of any one group do not skew the results within an experimental condition. Given that most classes are limited in number of students, dividing the class into smaller and more numerous discussion groups should enable researchers to test each condition using multiple discussion groups. Studies will also need to be conducted by varying male-to-female ratios. Furthermore, students will need more thorough instructions and training to reduce errors when assigning labels to messages. Finally, the patterns of interactions identified with the methods used in this study must ultimately be tested to determine how specific patterns on interaction contribute to group performance, group decision making, and learning outcomes.

Another area for future study is to determine to what extent message constraints and message labels can be used to support future research in CMC. There is the possibility that message constraints and message labels could alleviate some of the problems in parsing messages into discrete units of analysis, coding and assigning the units to functional categories, and analyzing the sequential patterns between messages and responses. The procedures used to implement message constraints and labels in this study will require further development before this technique can be used on a larger scale and used as a process-oriented approach to studying the factors that influence group interaction and the link between specific types of interactions and learning outcomes.

6. Instructional implications

The instructional benefits of using message labels remain in question due to the findings reported in this study. The findings serve to underscore the need to apply particular caution when using message labels to scaffold argumentation in online discussions given that the negative effect

sizes observed in this study were fairly substantial, although not necessarily conclusive. Nevertheless, the techniques of message labeling must be further developed, refined, and examined as a potential tool to assist both instructors and students in monitoring, developing and sustaining more critical discussions.

The findings in this study suggested a number of possible solutions to alleviate some of the potential problems with message labels: use more neutral labels to identify critiques like “question”, and “but”, and present a larger set of message categories and labels to make explicit and encourage multiple approaches to challenging arguments, such as testing for accuracy, validity, plausibility, and relevance. If future studies were to show that labels can improve argumentation under certain circumstances (e.g., all-male or mostly male discussions), the problems that remain (e.g., the way females might respond to certain labels) could be compensated by introducing additional strategies and requirements, such as designating students to play the role of devil’s advocate, explicitly require students to post a certain number of critiques and replies to critiques, balance the male–female ratio when possible, instruct and model the desired interactions to students, and set rules to prevent students from posting excessive and sometime redundant arguments so that more resources can be directed to challenging and analyzing individual arguments.

If the techniques of message constraints and labeling can be refined to a state where students can classify and label messages with high accuracy and make positive contributions to collaborative argumentation, these and other techniques demonstrated in this study may one day serve as a means to create discussion tools capable of providing real-time data and visual representations of group processes to facilitate online collaboration – particularly for highly structured tasks like argumentation. At the same time, such tools will also enable researchers to collect larger corpus of data across larger populations to thoroughly examine how different factors help to elicit interactions that produce critical discussions and how critical discussions help to improve group decision-making and problem solving.

References

- Bakeman, R., & Brownlee, J. (1982). Social rules governing object conflicts in toddlers and preschoolers. In K. H. Rubin & H. S. Ross (Eds.), *Peer relationships and social skills in childhood* (pp. 99–111). New York: Springer.
- Bakeman, R., & Gottman, J. (1997). *Observing Interaction: An introduction to sequential analysis*. Cambridge: University Press.
- Baker, M. (1999). Argumentation and constructive interaction. In P. Coirier & J. Andriessen (Eds.), *Foundations of argumentative text processing* (pp. 179–202). Amsterdam: Amsterdam University Press.
- Bakhtin, M. (1981). In M. Holquist (Ed.), *The dialogic imagination*. Austin: University of Texas Press.
- Beers, P. J., Boshuizen, E., & Kirschner, P. (2004). Computer support for knowledge construction in collaborative learning environments. Paper presented at the *American educational research association conference*, San Diego, CA.
- Buckingham-Shum, S. J., MacLean, A., Bellotti, V. M. E., & Jammond, N. V. (1997). Graphical argumentation and design cognition. *Human–Computer Interaction*, 12, 267–300.
- Butler, D., & Geis, F. L. (1990). Nonverbal affect responses to male and female leaders: implications for leadership evaluations. *Journal of Personality and Social Psychology*, 59, 48–59.
- Cho, K., & Jonassen, D. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research & Development*, 50(3), 5–22.
- Collins, C., & Collins, S. (1996). *The Internet as a tool*. (ERICK: ED 398 883).
- Derry, S. J., Levin, J. R., Osana, H. P., et al. (2000). Section on teaching, learning, and human development – Fostering students’ statistical and scientific thinking: lesson learned from an innovative college course. *American Educational Research Journal*, 37(3), 747–775.

- 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.
- Garrison, R. (2000). Theoretical challenges for distance education in the 21st century: a shift from structural to transactional issues. *International Review of Research in Open and Distance Learning*, 1(1), 1–17.
- Gottman, J. M. (1979). *Marital interaction: Experimental investigations*. New York: Academic Press.
- Gunawardena, C. N., Lowe, M. A., & Anderson, T. (1997). Analysis of a 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(4), 397–431.
- Hewitt, J., & Teplov, C. (1999). An analysis of growth patterns in computer conferencing threads. In *Proceedings of the computer support for collaborative learning (CSCL) 1999 conference*, Stanford University, Palo Alto, CA. Available from <http://www.ciltkn.org/csc199>. Retrieved June 01.06.04.
- Hiltz, S., Turoff, M., & Johnson, K. (1986). Experiments in group decision making: disinhibition, deindividuation, and group process in pen name and real name computer conferences. *Decision Support Systems*, 5, 217–232.
- Hollingshead, A. B. (1996). Information suppression and status persistence in group decision making: the effects of communication media. *Human Communication Research*, 23, 193–219.
- Jeong, A. (2003a). Sequential analysis of group interaction and critical thinking in threaded discussions. *The American Journal of Distance Education*, 17(1), 25–43.
- Jeong, A. (2003b). Gender interactions in on-line debates: look who's arguing with whom. Paper presented at the *Association for educational communications and technology conference*, Dallas, TX.
- Jeong, A. (2003c). The effects of message-reply and time-based structures on group interactions and critical thinking in asynchronous online discussions. Paper presented at the *association of educational communication & technology conference*, Anaheim, CA.
- Jeong, A. (2004). The combined effects of response time and message content on group interactions in computer-supported collaborative argumentation. *Journal of Distance Education*, 19(1).
- Johnson, D., & Johnson, R. (1992). *Creative controversy: Intellectual challenge in the classroom*. Edina, MN: Interaction Book Company.
- Jonassen, D., & Remidez, H. (2002). Mapping alternative discourse structures onto computer conference. Paper presented at *computer support for collaborative learning 2002 conference: Foundations for a CSCL community*, Boulder, CO. Available from <http://newmedia.colorado.edu/csc1/12.html>. Retrieved 20.02.02.
- Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtin's Contribution to Understanding Learning in Settings of Collaboration. In *Proceedings of the computer support for collaborative learning (CSCL) 1999 conference*.
- Leinonen, T., Virtanen, O., & Hakkarainen, K. (2002). Collaborative discovering of key ideas in knowledge building. In *Proceedings of the computer support for collaborative learning 2002 conference*, Boulder, CO. Available from <http://fle3.uijah.fi>. Retrieved 19.05.04.
- Lemus, D., Seibold, D., Flanagan, A., & Metzger, M. (2004). Argument and decision making in computer-mediated groups. *Journal of Communication*, 54(2), 302–320.
- 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.
- Lipman, M. (1991). *Thinking in education*. Cambridge: Cambridge University Press.
- McAlister, S. R. (2001). Argumentation and a design for learning. (CALRG Report No.197). Available from <http://iet.open.ac.uk/pp/s.r.mcalister/personal/197.pdf>. Retrieved 03.03.04.
- McAlister, S. (2003). Assessing good argumentation. Available from <http://iet.open.ac.uk/pp/s.r.mcalister/personal/AssessingGEA.htm>. Retrieved 10.04.04.
- Newman, W. R. (1995). The psychology of the new media. *Educom Review*, 30, 48–54.
- Olson, G., Herbsleb, J., & Rueter, H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. *Human-Computer Interaction*, 9(3/4), 427–472.
- Poole, M. S., Holmes, M., Watzon, R., & DeSanctis, G. (1993). Group decision support systems and group communication: a comparison of decision making in computer-supported and nonsupported groups. *Communication Research*, 20(2), 176–213.

- Ravits, J. (1997). An ISD model for building online communities: furthering the dialogue. In O. Abel, N. Maushak, & K. Wright (Eds.), *Proceedings for the 1997 national convention of the association for educational communications and technology* (pp. 297–307).
- Rourke, A., 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.
- Saba, F., & Shearer, R. (1994). Verifying key theoretical concepts in a dynamic model of distance education. *The American Journal of Distance Education*, 8(1), 6–59.
- Sappo, T., & Mononen-Aaltonen, M. (1998). Developing dialogic communication culture in media education: integrating dialogism and technology. *Media Education Publication*, 7.
- Schmidt, K. (1999). Of maps and scripts: the status of formal constructs in cooperative work. *Information and Software Technology*, 41, 319–329.
- Sloffer, S., Dueber, B., & Duffy, M. T. (1999). Using asynchronous conferencing to promote critical thinking: two implementations in higher education. Paper presented at the 32nd conference on system sciences, Maui, Hawaii, January 5–8.
- Sproull, L., & Keisler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, USA: MIT Press.
- Stegmann, K., Weinberger, A., Fischer, F., & Mandl, H. (2004). Can computer-supported cooperation scripts facilitate argumentative knowledge construction? Paper presented at the 2004 American educational research association conference, San Diego, CA. Available from <http://home.emp.paed.uni-muenchen.de/~weinberg/download>. Retrieved 24.05.04.
- Stern, D. N. (1974). Mother and infant at play: the dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–213). New York: Wiley.
- Suthers, D. (1998). Representations for scaffolding collaborative inquiry on ill-structured problems. Paper presented at the 1998 American educational research association conference, San Diego, CA.
- Suthers, D. D., & Hundhausen, C. D. (2001). Learning by constructing collaborative representations: an empirical comparison of three alternatives. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 577–584). Maastricht, Netherlands: Universiteit Maastricht.
- Suthers, D., Hundhausen, C., & Girardeau, L. (2003). The roles of representations in face-to-face and online computer supported collaborative learning. *Computers & Education*, 41(4), 335–351.
- Tan, S.C. (2000). Supporting collaborative problem solving through computer-supported collaborative argumentation. Unpublished doctoral dissertation. The Pennsylvania State University, Pennsylvania.
- Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: Ballentine Books.
- Tisdell, E. J. (1993). Interlocking systems of power, privilege, and oppression in adult higher education classes. *Adult Education Quarterly*, 43(4), 204–226.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: University Press.
- Turoff, M., & Hiltz, R. S. (1998). Superconnectivity. *CACM*, 41(7).
- Vanfossen, B. (1996). Gender differences in communication. *Institute for teaching and research on women*. Towson, MD: Towson University. Available from: <http://www.towson.edu/~vanfoss/wmcomm.htm>. Retrieved 14.03.02.
- Veerman, A. L., Andriessen, J. E. B., & Kansellaar, G. (1999). Collaborative learning through computer-mediated argumentation. Paper presented at the conference on computer supported collaborative learning (CSCL'99), San Francisco, CA.
- Ward, S. C., & Tiessen, E. L. (1997). Adding educational value to the Web: active learning with a live page. *Educational Technology*(September–October), 22–30.
- Weinberger, A., Fischer, F., & Mandl, H. (2001). *Scripts and scaffolds in problem-based computer supported collaborative learning environments: fostering participation and transfer* (Research report No. 144). Ludwig-Maximilians University, Institute for Empirical Pedagogy and Pedagogical Psychology, Munich. Available from <http://home.emp.paed.uni-muenchen.de/~weinberg/download>. Retrieved 24.05.04.
- Weinberger, A., Fischer, F., & Mandl, H. (2004). Knowledge convergence in computer-mediated learning environments: effects of collaboration scripts. Paper presented at the American educational research association conference, San Diego, CA.
- Wiley, J., & Voss, J. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91, 301–311.