Effects of Collaboration and Argumentation on Learning from Web Pages

Jennifer Wiley

Jeannine Bailey

University of Illinois at Chicago

Many recently developed learning environments incorporate collaborative learning as well as the use of technology in their design. However, at least within the social psychology literature that has investigated groups as information processors, studies have found that groups of people working together are frequently less effective than individuals working alone, or than the combined efforts of an equal number of individuals working alone (Diehl & Stroebe, 1987; Gigone & Hastie, 1997; Hastie, 1986; Hill, 1982). On the other hand, among studies of collaborative learning in educational contexts, advantages of the performance of a group over individuals have been reported (Hertz-Lazarowitz & Miller, 1995; Johnson & Johnson, 1989; Webb & Palincsar, 1996). The focus of this chapter is to begin to explore what may be responsible for the discrepancy between the findings of these two literatures, the kinds of processes that may underlie gains and losses in group contexts, and finally, the use of technology to promote successful collaborative problem solving and reasoning.

Factors Determining Process Loss

Intact groups are less productive, less creative, and often biased or unduly-influenced by one member (Diehl & Stroebe, 1987; Gigone & Hastie, 1997; Hastie, 1986; Hill, 1982). The term process loss was coined by Steiner (1972) to describe the loss in productivity that occurs when individuals
must coordinate their efforts in a group. On the other hand, when there is a synergetic or value-added effect observed among individuals working as a group versus individuals working alone, this has been termed process gain (Steiner, 1972).

The social psychology literature has identified a number of factors that seem to contribute to process loss. For this discussion, we have grouped their observations into four aspects of group problem solving: motivation and effort, evaluation and conflict, task coordination, and use of resources. Proponents of collaborative learning who have suggested that working in groups may contribute to better learning cite these same aspects of group settings as potential reasons for learning gains (e.g., Webb & Palincsar, 1996). Thus, it seems that each factor may have opposing effects on performance in different contexts. The goal is to figure out which contexts lead to positive outcomes, (i.e., process gain), and which contexts result in negative outcomes, (i.e., process loss).

Changes in Motivation and Effort

Since the late 19th century, researchers have recognized that the mere presence of others can be motivating, whether it is turning fishing reels (Triplett, 1898), or generating word associations (Allport, 1920). When observers are present, especially when they may be in a position to evaluate performance, individuals become more productive. This has been termed social facilitation (Allport, 1924).

On the other hand, classic studies have also demonstrated that as soon as more than one person is responsible for a task, there is a risk that not all will participate fully (Ringelmann, 1913). When an individual is solely responsible for a task then she or he must be motivated or involved for it to be accomplished. As soon as more than one person is responsible for a task, social loafing may occur (Kerr & Bruun, 1983). This term refers to the fact that as more people are responsible for performing a task, each individual feels less responsibility for its actual completion, and as a result, each may contribute less than he or she would individually. Process loss has frequently been attributed to losses in motivation, a lowered sense of responsibility, and less effort by each individual (see Sheppard, 1993). Group problem solvers contribute fewer guesses in a group than alone (Mullen, Johnson, & Salas, 1991; Taylor, Berry, & Block, 1958), and group decision makers feel less investment in (and responsibility for) their decisions (Gigone & Hastie, 1997).

Potential for Evaluation and Conflict

One potential explanation for the loss of motivation or effort is anxiety or fear about being evaluated by others. The presence of others introduces an element of evaluation and conflict, not usually present when people act
alone. This can have advantages, and theoretically could improve the quality of the group's contribution. For example, others may detect errors and provide immediate feedback to any individual in the group. On the other hand, the potential for being evaluated can also have an inhibiting effect, and working with others can cause evaluation apprehension, causing poorer performance and the generation of fewer or less creative ideas (Allport, 1924; Camacho & Paulus, 1995).

The potential for conflict and disagreement may exist as soon as a second person is added to the task. Further, the presence of at least three people in a group may make evaluation, conflict, and negotiation of positions even more likely, as the third person establishes majority and minority stances. Evaluation becomes especially important, as groups need to choose which information to consider and which opinion which to give more weight. A single person has only his or her own opinion. Multiple people need to consider other perspectives and determine which perspective is correct. Problem-solving studies have demonstrated that in unsuccessful groups, the minority positions are not attended to, or the opinion of the most talkative member is accepted (Maier & Solem, 1952; Thomas & Fink, 1961, 1963).

There is an additional level of social factors that can seriously impact the evaluative activity of a group. Perceptions of inequality in power, status, and knowledge about the topic all impact who contributes in a group and whose opinions are attended to in a discussion. Further, there is often a strong push toward conformity and avoidance of conflict in group settings, which can result in convergence on nonoptimal solutions, or groupthink.

Task Coordination

One person acting alone does not need to reflect on the task or the process for solution. A group must make these processes explicit. To the extent that the group fails to reach an understanding of the task, or a process for coordinating actions, there will be wasted effort and resources. However, the potential exists for the group to engage in more reflective and mindful problem representation than an individual (Moreland & Levine, 1992).

Studies have directly observed that some process loss occurs due to coordination difficulties by comparing intact groups (people who have worked together) with ad hoc groups (individuals put together for the first time). Ad hoc groups need some time together to get past organizational issues. When ad hoc groups are given an extended period of time, deficits go away (Anderson, 1961; Watson, 1928). On the other hand, intact groups have their own set of additional problems. Intact groups have coordination schemes in place, but may also have biases about which group members are usually the best source of information. In a classic study, Torrance (1954) nicely demonstrated that B-27 bomber teams disproportionately weighted the opinion of the pilot on problem-solving tasks completely unrelated to
flying, even though that weighting was irrelevant in the new problem-solving context.

Use of Resources

One of the most obvious reasons for expecting two heads to be better than one is due to the ability of group members to pool resources. The group has the advantage of multiple frames with which to select relevant or important information to attend to, multiple long-term memory stores from which to retrieve relevant knowledge, and multiple buffers that may allow for more elaborate processing. The presence of collaborators increases the amount of information that can be brought to the table, increases the probability that any particular person will be aware of a critical piece of information, and increases the amount of information that can be considered simultaneously. Individuals can also prime others or remind them of relevant facts from long-term memory.

From an information processing approach, the presence of multiple perspectives, stores, and buffers should improve performance on complex learning tasks, which may tax the abilities of a single individual. However, this benefit from collaborators can only be the case when all people actually participate and freely share the unique perspective that they have. Unfortunately, studies from the social psychology literature suggest that this rarely happens (Hinsz, Tindale, & Vollrath, 1997; McGrath, 1984). Group members tend to offer more shared than unique information into discussion (Larson, Foster-Fishman & Keys, 1994; Stasser & Titus, 1985) and as noted briefly earlier, individuals with unique information are often ignored by the rest of the group. For the minority position to have an influence, the person needs to be confident, talkative, or able to demonstrate that his or her solution is correct, or there needs to be a leader who is responsible for making sure minority positions are heard (Hastie, 1986; Maier, 1967; Oxley, Dzindolet, & Paulus, 1996). Alternatively, expert roles can be assigned to encourage the sharing of unique information (Hollingshead, 1996; Stewart & Stasser, 1995). In these cases, role assignment can lead to some increases in group performance.

Even more discouraging are the most recent findings in this literature, which suggest that group work can actually increase the processing load on individuals. Interacting in groups increases the amount of information available immediately, but this means an increase in the amount of information that needs to be processed. Working with others adds information that needs to be attended to (individuals need to consider the task at hand, their own thoughts, and other people’s thoughts and opinions). In addition, the individual has less time to state his or her own thoughts out loud, and others’ contributions may interrupt the individual’s own processing, knocking them off their train of thought. Further, more time may be spent on off-task topics (Dugosh, Paulus, Roland, & Yang, 2000). In the end, individuals
may be more burdened and enjoy less intact cognitive processing than when working alone. Studies have directly tested this notion of interrupted cognitive processing among groups, which has been called production blocking and support an account of process loss as a function of increased disruption in group situations (Diehl & Stroebe, 1987; Stroebe & Diehl, 1994).

Thus, the social psychology literature has demonstrated that poorer group performance can result due to many possible factors. As a result of changes in motivation and effort, fear of evaluation, and coordination and communication burdens, groups are often observed to perform more poorly than individuals. However, the social group problem-solving literature has largely ignored classroom or educational contexts (e.g., Gigone & Hastie, 1997); the studies in this literature generally do not have authentic or naturalistic classroom tasks, and they tend to study short-lived interactions of ad hoc groups. Nevertheless, studies of group information processing have identified several ways in which working in groups may lead to losses in performance. As educational and cognitive researchers delve deeper into explaining successful collaborative learning, a valuable contribution may be made by addressing the factors that have been seen to cause process loss outside learning contexts, and discovering how successful learning environments may be circumventing these potential losses.

Educational Research on Collaborative Learning

Researchers in educational psychology have viewed collaborative learning from a number of perspectives. Through interaction with others, learners jointly construct or instantiate knowledge structures for themselves. Some theoretical advantages of working in groups are based in mechanisms of cognitive development, where learning is prompted by conflict with others (Perret-Clement, 1980; Piaget, 1932), internalization of social processes from working with others (Vygotsky, 1978), and modeling of or for others (Bandura, 1986; Brown & Palincsar, 1989). Other approaches hinge positive outcomes on the creation of interdependence among group members (Johnson & Johnson, 1989; Slavin, 1987). These perspectives have guided the design of collaborative learning environments that support interaction among peers.

Most collaborative learning studies that have demonstrated positive educational outcomes have investigated how better learning may result from different kinds of learning activities. Most often comparisons are between structured and unstructured collaborative learning, and advantages are found for supportive or structured collaboration. For example, Coleman (1998) investigated the development of student understanding of photosynthesis by either having students interact in scaffolded explanation groups, or in nonscaffolded groups. The scaffolding prompted students to give evidence, build on others' ideas and to try to understand others' positions. With these prompts, the scaffolded collaboration groups acquired a better
understanding than the nonscaffolded groups. In another line of research, King (1990) found that instruction in reciprocal questioning strategies is more effective for learning than class discussion or reviewing material (see also Fantuzzo, Riggio, Connolly, & Dimeff, 1989).

Alternatively, researchers have looked at the interactions within small groups to determine what aspects of collaboration may actually contribute to learning gains. Studies in this vein include work by Webb, Troper, and Fall (1995), which find that student-generated explanations are related to better learning, both in the students who generate the explanations and in students who hear the explanations, as long as the hearers actually apply the explanation on their own after they hear it.

Only a few studies have directly compared learning or transfer of learned material among students learning in collaborative groups to individual learning using similarly structured tasks or environments. O'Donnell and Dansereau have been responsible for a number of studies in the literature that explore the effectiveness of interventions developed for collaborative learning settings, in both collaborative and individual learning contexts. Examples from their findings include that students who unexpectedly review a lesson in pairs perform better than individuals who reviewed alone, or dyads who expected to collaborate (O'Donnell & Dansereau, 1993). In another study, dyads learned more successfully when both students in a pair were simultaneously given a script (a learning organizer that scaffolded their activity), as opposed to individuals with or without scripts, and dyads who were not given scripts (O'Donnell, Dansereau, Hall, & Skaggs, 1990).

In a comparison of individual versus collaborative mathematical problem solving, Vye, Goldman, Voss, Hmelo, and Williams (1997) found that dyads explored more solution paths and had more coherent reasoning about the problem than did individual problem solvers. More recently, Barron (2000) found that collaborative problem solving among triads can also lead to improved mathematical problem solving. Triads outperformed individual problem solvers on an initial complex mathematical problem. More important, students who had collaborated on the initial problem then outperformed students who had worked individually on transfer problems. In general, the literature on collaborative learning tends to lack comparisons among individuals learning in a single context versus a matched collaborative context, and rarely are there tests of whether or not individual learners can transfer their learning to new material following the collaboration. The demonstrations just described are thus critical pieces of evidence that at least in some cases, collaborative learning contexts may be circumventing or capitalizing on the aspects of group interaction that might otherwise lead to losses.
12. LEARNING FROM WEB PAGES

Why So Little Process Loss in Studies of Collaborative Learning?

If we look for conditions that seem to maximize the potential of collaborative learning, we might expect process gain to be found in contexts that address the four areas of potential losses outlined previously: loss due to motivation, fear of evaluation, coordination, and communication burdens. Following from that logic, conditions that may make process gain most likely could be contexts that require the integrated effort of all members, contexts that provide a basis for evaluation, contexts that scaffold coordination and planning, and contexts that structure communication.

Motivation Through Interdependence

Several researchers have commented that successful collaboration directly depends on the extent to which the task can be completed without contributions of all the members (Steiner, 1972). In a similar vein, Hertz-Lazarowitz and Miller (1995) suggested that successful collaboration only occurs when the task requires the integration of the thinking of all members. Only on these interdependent tasks, where students must cooperate for the goal to be achieved, will all students be motivated to participate (Johnson & Johnson, 1989). It is these kinds of cooperative, collaborative tasks that are related to increases in student effort and drive, and that result in the findings that classrooms that employ more collaborative learning activities result in more student time on task and fewer absences (Slavin, 1987). Larson and Christensen (1993) also suggested the need for interdependence, true interaction, and complexity to see benefits. Tasks in the social psychology literature have rarely required this. The work of Johnson and Johnson (1989) suggests that studies in the social literature may be collapsing across cooperative and competitive groups. Only the former will promote better learning from groups.

Norms for Evaluation

Other research, such as the Coleman (1998) study described earlier, suggests that the potential for evaluation and conflict may only lead to higher levels of reasoning and learning when students are given norms for evaluation. When ground rules are set such that students must rely on a comparison of evidence as means to reconcile conflict, when they need to explain their position to others, and when they need to justify their opinions, only then might we find better learning in group contexts. Once these ground rules are established, then the generated questions, explanations, and proofs may provide important models for students to emulate and internalize.
Better Task Representation

Similarly, the need to discuss the task with others can also support better task definition, goal setting, and planning activities. When students are put in contexts with specific tasks, it may help them to plan their problem solving or learning, and they may get past coordination problems faster. Finally, contexts that structure interaction, and especially tasks that focus learners on important information, and encourage the contribution of prior knowledge and unique perspectives, will be important for effective group activity.

Supporting Collaborative Learning with Argumentation Tasks

Fulfilling these conditions may be the mechanism by which successful learning contexts can achieve performance gains from collaborative learning. This study will investigate whether positive outcomes in collaborative learning studies may be seen in educational contexts with several key elements: when small groups are given a specific learning task that requires full participation, provides norms for evaluation, and channels attention toward important information and communication toward explanatory behaviors.

The learning task that we think may fulfill these demands is an argumentation task. As such, this study investigates the effects of collaboration in the context of an argument writing task from multiple sources. In previous studies, it has been demonstrated that argumentation is a task that supports better understanding in students working alone (Wiley, 2001; Wiley & Voss, 1996, 1999). In these studies, learners were provided with several texts about a topic, presented in a Web site, and were asked to construct an argument of “What produced the significant changes in Ireland’s population between 1846 and 1850?” or “What produced the eruption of Mt. St. Helens in 1981?” When students were given an argument writing instruction, as opposed to a narrative writing or essay writing instruction, they wrote more causal and integrated essays, and performed better on outcome measures such as inference and analogy recognition tests, suggesting that argument writing led to better understanding of the subject matter. The advantage was especially pronounced when students learned from multiple sources, presented in a Web site, than when they learned from a single text. When students were put in a browser environment that explicitly supported the comparison and integration of multiple documents, and were given an argumentation task that also prompted integrative and constructive activity, they demonstrated better conceptual understanding of the texts as compared to more traditional textbook learning conditions (Wiley, 2001; Wiley & Voss, 1999).
There are several reasons why argumentation itself may be a powerful learning instruction (Voss & Wiley, 2000). Implicitly, the instruction to construct an argument may require that students develop their own personal thesis or theory. Additionally, students may also perceive the argument instruction as constructing an account that includes evidence to support or justify a position (as in the sense of a legal argument). The term argument may imply that a position needs to be defended to others, and may cause the learner to represent an external discourse. A final reason why the argument writing task may be essentially powerful is that in supporting a thesis, the learner must transform the presented information, and integrate it into a new coherent explanatory structure to fulfill the argument writing task. Especially the act of finding new connections across texts, to justify and explain in an argumentation task, may be critical for the better understanding achieved in this context (Wiley & Voss, 1999).

In many respects, the ways that argumentation tasks promote understanding appear similar to notions of how collaborative tasks are thought to facilitate learning. Further, it seems that argumentation tasks may be effective in collaborative learning contexts for reasons over and above the reasons why they may help individual learning. As a task, argumentation tasks may motivate learners by seeming to be personally relevant. Both learners may have an investment in developing an argument that they agree with. Second, argumentation may provide collaborative learners with an evaluation structure that may help them avoid many of the pitfalls of group problem solving. Giving groups a task that explicitly prompts the generation of hypotheses, and the need to support and justify those hypotheses may attenuate many issues of evaluation. Evaluation no longer runs as big a risk of being personal or alienating as it is in the best interest of all students to espouse a position that can be supported by evidence. Further, the use of claims and evidence can make solutions demonstrable or thinking observable to other students. This not only helps modeling, but is also consistent with the notion that groups are more effective when solutions can be demonstrated as correct. Third, argumentation is a specific task that can be reflected on and planned for. Students may be more likely to engage in planning activities with a specific task than with a vague one that may be fulfilled by any product. Finally, the argumentation task can be used to structure interaction or communication by keeping students focused on explanatory activity (attending to important information in the text, and using it to construct an explanation). Thus, the use of an argumentation task may not just reinforce the strengths of collaborative learning; a task like argumentation may in fact be necessary to reap the benefits of working in groups. In line with this perspective, it seems particularly fitting that Webb and Palincsar (1996) suggested that there is heightened interest among the collaborative learning community in situations that require elaboration, interpretation, explanation, and argumentation. These processes may specifically fulfill the needs of collaborative learners as information processors,
and allow them to support each other rather than impede each other’s learning.

The Present Study

Pairs of learners were observed as they navigated through Web sites related to volcanic eruptions, to write a summary or an argument of “What caused the eruption of Mt. St. Helens?” Pairs were asked to jointly compose the essay. Following the reading and writing phase, all students individually completed two learning outcome measures: an inference verification task and an analogy rating task that required the recognition of similar and dissimilar geological events.

The main questions of interest are these: What do successful collaborative learners do as they complete the reading and writing tasks? Does argumentation lead to better collaborative learning? If so, what behaviors can we identify among collaborative arguers that may be responsible for supporting understanding? To assess whether collaboration is impacting motivation, we look at time on task across conditions, and number of utterances. We also look at the perceived interdependence of the group, by looking at the number of questions they ask each other. To examine the extent of evaluative activity among pairs, we look at whether conflict occurs, whether students evaluate the sources or each other, and whether they evaluate their own understanding. In terms of coordination of activity, we look at whether time is spent explicitly planning, discussing, or coordinating the execution of the reading and writing tasks. Finally, to examine communication-related effects, we examine how much information is exchanged, and what kind of information is stated in each condition.

Although the study as presented is exploratory, it is hoped that the interaction of collaborative and argumentative task environments may provide for better understanding of the subject matter, and the goal is to be in a position to describe in detail behaviors and patterns of interactions that may be contributing to success in this learning context.

METHOD

Participants

Four pairs of undergraduates participated in this study. The undergraduates were recruited from the University of Illinois at Chicago participant pool and received course credit for their participation. Pairs were not familiar with each other before engaging in this task. Two pairs were given each
writing task. Of the two pairs given the summary writing task, one had two men and the other had two women. Both pairs given the argument task had one man and one woman. This is an unfortunate chance occurrence, and it should be noted that any differences in group dynamics between the two conditions could be attributed to the differences in dyad composition.

Materials

The reading material for this study was made to mimic the output of a Google search using these keywords: causes volcanic eruptions. There were two output pages, each with links to five Web sites. On the first page, the sources listed were a NASA page on volcanoes; an astrology site “Blast from the Past” on how Mt. St. Helens erupted because of the alignment of planets; a United States Geological Survey site on earthquakes, volcanoes, and plate tectonics; a Scientific American site where an expert explains the causes of volcanic eruptions; and a page from Vocanolive.com on how tides relate to eruptions. The second Google search page had links to five more sources: a commercial site for the Iben Browning Newsletter, written by a consultant who predicts earthquakes and volcanic eruptions; a PBS site for the series Savage Earth; a PBS site on plate tectonics for the series A Science Odyssey; a commercial site for the Cook Internal Propulsion Engine that claims that oil drilling is the cause of earthquakes and volcanic eruptions; and a site from CPB/Annenberg on volcanoes. Most sites contained many pages. The amount of information potentially available to the students was intentionally more than could be read in an hour. This was meant to necessitate decision making and negotiation among the pairs, as well as increase the chances of finding an advantage for multiple learners, because there was clearly more information than a single information processor could handle. The inclusion of a range of sources that varied in their reliability was also intentional, and was intended to prompt evaluative comments on the part of the learners.

Procedure

Participants were tested in pairs in front of a single computer. The experimenter was present in the room for the entire session. Sessions lasted 1 hour. On arrival, students were informed of the video recording and signed a consent form. They were then each given a pretest and told to complete the five short-answer questions to the best of their ability (e.g., What is a volcano? Why do volcanoes occur?). They were then told that they would be writing either an argument or a summary about the cause of the eruption of Mt. St. Helens. To help them with their task, we told them that we had done a Google search for them using the keywords of causes volcanic eruptions. The results of the search were on the screen in front of them and con-
sisted of two pages of output with a total of 10 links that they could use to write their essay. They were also told that they were to work together on this task and that they would have around 30 minutes to both read the web pages and write their essay. They were given paper to write their essay and extra paper to take notes if they wanted. The experimenter stayed in the room and took notes as they worked. With 10 minutes left, the experimenter encouraged them to finish their essay. On completion, the video camera was turned off and they completed a posttest booklet individually and without the computer.

The posttest booklets consisted of three parts: the inference task, the analogy task, and the final survey page. The inference task consisted of 25 items. Five items were distractors that were not explicitly addressed in the reading material. Ten items were directly related to central concepts in the reading. The remaining 10 items were related to less central details in the texts.

The analogy task required students to judge the similarity of several short scenarios describing potentially analogous events. The task consisted of five items. Students were asked to rate the similarity of the cause of each event to the cause of the Mt. St. Helens eruption on a 10-point scale with 10 meaning very similar. The first item was a practice item on Hawaiian volcanoes. Two items concerned events that were not causally similar to the Mt. St. Helens eruption: California wildfires and the New Madrid, Missouri, earthquakes. These items are included in the dissimilar events rating measure later. Low scores on these items indicate understanding. Two items concerned causally similar events: the Kobe earthquake and the eruption of Mt. Pinatubo. These items are included in the similar events rating measure later. Higher scores on this measure indicate understanding. Finally, a discrimination index was computed by subtracting the average rating for dissimilar items from the average rating for similar items. Higher positive scores represent better discrimination of similarity.

The last page of the booklet was used to gather other information regarding the participants. Descriptive information included their age, gender, and intended major. They were also asked to define an argument and a summary, what the major difference was between them, and whether or not they liked learning with a partner and why.

Coding

Several categories were established a priori based on the literature and our predictions. We were interested in the number of utterances, number of questions (regardless of content), number of interactions, number of evaluative statements, number of planning-related statements, and kind of information shared. In reviewing our protocols, we operationalized these categories and developed the detailed coding scheme for each utterance pre-
sentenced in Table 12.1. Two independent raters coded all four protocols with category agreement on 89% of utterances.

**TABLE 12.1**

Coding Categories

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>Propositional utterance units</td>
<td>Each utterance broken into propositions</td>
<td>The planets caused the volcano?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ready for page 2?</td>
</tr>
<tr>
<td>Q</td>
<td>Questions</td>
<td>Questions double-coded with categories</td>
<td>Are you done reading?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yeah.</td>
</tr>
<tr>
<td>P</td>
<td>Polite expressions</td>
<td>Turn-taking prompts, clarifications, or acknowledgments</td>
<td>I don't think we have to read that.</td>
</tr>
<tr>
<td>O</td>
<td>Off-task behavior</td>
<td>Anything not related to reading or writing</td>
<td>That's real good for my self-esteem.</td>
</tr>
<tr>
<td>E</td>
<td>Evaluation</td>
<td>Value judgments</td>
<td>This stuff is just garbage.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>It's not telling us why...</td>
</tr>
<tr>
<td>M</td>
<td>Metacognition</td>
<td>Assessment of understanding</td>
<td>I'm not sure I get this.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Do you understand this?</td>
</tr>
<tr>
<td>R</td>
<td>Planning of reading</td>
<td>Comment on reading goals or process</td>
<td>I don't think we have to read that.</td>
</tr>
<tr>
<td>T</td>
<td>Comment about writing task</td>
<td>Comment on task process or demands</td>
<td>I think this is long enough.</td>
</tr>
<tr>
<td>W</td>
<td>Comment about content</td>
<td>Suggestion of content for writing</td>
<td>I guess we need to lead into this subduction part.</td>
</tr>
<tr>
<td>C</td>
<td>Claim</td>
<td>A theory or summary statement of a claim</td>
<td>The planets caused the volcano?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>It's how every volcano erupts.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This guy's talking about Turkey...</td>
</tr>
<tr>
<td>S</td>
<td>Evidence from source</td>
<td>Statement of a piece of information</td>
<td>A 5.1 earthquake opened a vent.</td>
</tr>
<tr>
<td>K</td>
<td>Evidence from knowledge</td>
<td>Statement of a piece of knowledge</td>
<td>I think I remember this.</td>
</tr>
</tbody>
</table>
RESULTS

Prior Knowledge Assessment

The prior knowledge of each participant was assessed individually in a pre-test with these five questions: What is a volcano? Why do volcanoes occur? How do volcanoes erupt? Where are volcanoes located? What kinds of volcanoes are there? The conceptual model possessed by each student was categorized using the answer to these questions and a taxonomy of volcanic understanding developed by Hemmerich and Wiley (2002). Incorrect models of volcanic eruptions, such as that they are caused by the weather or gods, are coded as Level 0. Partially correct models that only include a single local cause, of either plate movement, heat flows, or pressurized magma, are coded as Level 1. Models that contain more than one correct cause are coded as Level 2. A model that integrates plate movement with heat or pressure factors represents a mature understanding of volcanic activity and is coded as Level 3. Of our four pairs, one pair in each condition had no prior understanding of volcanoes (i.e., both members started at Level 0). The other two pairs each had one member with some understanding of the role of earth movement or plate tectonics (at Level 1), and the other member with a Level 0 understanding. Thus, in some sense, our pairs were matched for prior conceptual understanding about volcanoes.

Essay-Based Pair Learning Outcomes

The essays that the students wrote together were again coded in relation to the taxonomy of volcanic understanding. In their essay, Pair 1 asserted that the alignment of the planets caused the eruption of Mt. St. Helens. This is an incorrect model, and was coded as Level 0. Pair 2 wrote in their essay that Mt. St. Helens erupted because the plates at Mt. St. Helens crunched together, that plates move because of heat flow, and that an earthquake opened a vent that caused the eruption of pent up gases and magma. Thus in their essay, they recognized the relation between plate movement and heat flow, and that these two factors and pressure relate to volcanic eruptions. Because the explanations of plate movement and heat flow are integrated, the essay is coded as demonstrating a Level 3 model of volcanic activity. This is a more complex model than either participant held before the learning activity. Pair 3 wrote in their essay that plate tectonics causes rifts and ridges in the Earth’s crust. They also asserted that volcanoes are formed when hot rocks move and fill chambers in the Earth’s crust. This is a model that contains both elements of movement and heat, but because the heat is not directly related to the movement of plates or to the melting of the plate, it is coded as a Level 2 model. The final pair, Pair 4, wrote an essay that
included the ideas that plate movements cause the formation of magma in subduction zones, and that this magma is especially stiff, which leads to the tremendous buildup of internal gases. By mentioning both plate movement and pressure, and linking them through the formation of viscous magma, these students created an integrated model of volcanic activity, and the essay is coded as Level 3. Unfortunately, the students then went on to say that "another contribution to volcanic eruptions is the drilling of too much oil." This pet theory of the inventor of a perpetual motion engine gets assimilated into an otherwise good account of the causes of the eruption of Mt. St. Helens.

Individual Learning Outcomes

To assess the amount of learning in each partner as a result of the reading and writing tasks, we examined each student's individual performance on inference and event similarity posttests. Average performance was better on the inference and the dissimilar events rating task for pairs who had an argument task than pairs who had a summary task. As presented in Table 12.2, the outcome measures were more systematic in the argument condition than in the summary condition. Students in the argument condition tended to recognize more critical inferences \((M = 6.75, SD = .75)\) than students in the summary condition \((M = 4.5, SD = 2.08)\). No differences were seen in the Similar Event ratings \((Argument M = 5.13, SD = .85; Summary M = 4.75, SD 1.19)\). However, the argument participants were more likely to recognize the dissimilarity of the two scenarios \((M = 3.25, SD = 1.5)\) than were participants in the summary condition \((M = 5.75, SD = 1.5)\). They also had better discrimination between similar and dissimilar events. These data suggest generally positive learning outcomes in the argument condition, but also for one pair in the summary condition. The question of interest is whether we can observe any discourse patterns that may be related to the development of understanding, which in turn could support better learning.

Dialogue-Based Analyses

To answer these questions, we coded the interactions of the pairs as they read and wrote their essays. The results of our protocol analysis are presented in Table 12.3. There was very little off-task behavior in the protocols, so this category is not included. No one spoke about the weather or classes or anything outside the task. Although this is quite a small sample, there are a few patterns in the data that can be identified. In general, the two pairs in the argumentation condition seemed to interact more than the summary pairs. Both argumentation pairs had a high number of total utterances, asked a good number of questions, and made many polite remarks. More important, the three dyads (Pairs 2, 3, and 4) that showed positive
learning outcomes all spent more time talking specifically about essay content and negotiating the writing process than did the nonlearning pair (Pair 1).

TABLE 12.2
Learning Outcomes by Pair and Individual

<table>
<thead>
<tr>
<th>Task</th>
<th>Summary</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3</td>
<td>4 5 6 7 8</td>
</tr>
<tr>
<td>Pair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>F F M M</td>
<td>F F M</td>
</tr>
<tr>
<td>Sex</td>
<td>0 0 1 1</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Pretest score</td>
<td>3 3 2 2</td>
<td>3 3</td>
</tr>
<tr>
<td>Essay score</td>
<td>4 7 2 5</td>
<td>7 6 6</td>
</tr>
<tr>
<td>Inference score</td>
<td>5 3 5.5 5.5</td>
<td>5 6 4</td>
</tr>
<tr>
<td>Similar event rating</td>
<td>7 5 7 4 3.5 4</td>
<td>4.5</td>
</tr>
<tr>
<td>Dissimilar event rating</td>
<td>-2 -2 -1.5 1.5 4.5</td>
<td>1.5 2 2 2</td>
</tr>
</tbody>
</table>

These preliminary results suggest that collaborative learning will be most successful when students interact well and negotiate their understanding of the subject matter. When learning partners do not engage each other, then an opportunity may be lost. This comparison suggests that argumentation tasks may support collaborative learning through providing students with a specific task and medium for interaction about the task and content. More successful groups, regardless of the writing task, engaged in more coordination of understanding about the topic. Interestingly, there were not a large number of evaluative statements in either writing condition, and those that were made did not seem to contribute much to learning (this was true considering the number of evaluations made by either member of the pair). Hence, it does not seem that the argumentation task necessarily promotes better evaluation of content.

In addition to these attempts to code and quantify the interactions among the pairs, examples of the actual dialogue better illustrate the dynamics of each dyad. For example, here is an exchange from Pair 1:
Betty reads from NASA site: Quieter volcanoes, like Iceland and Hawaii, are found mostly where plates are coming apart or in the middle of a plate. Veronica: Go on to the next one. (Goes to Blast From the Past astrology site). Betty reads from site “Earthquakes and volcanoes have long been associated with planets.”

This exchange is typical of the interaction between these two women. When they did speak, they were reading information from the site to each other. They did not negotiate either the writing of the essay or their understanding of the material. Here is an example of Pair 2:

Archie (clicks on A Science Odyssey Web site): Let’s just check on these other ones really fast and see what they are. Ralph: Yeah, so this says the movement of plates is what causes earthquakes and volcanoes and you could maybe write this down—what causes plates to move is unknown. Let me write. Okay, one theory is heat convection from the earth’s inner core. Archie: Okay.

In this pair, the second man (Ralph) takes the lead, produces almost all the claims and text on his own, and eventually takes over the actual writing. There is limited discussion, however, about the subject matter and content of the essay. In the next example, Pair 3 engages in more negotiation:

Frank: OK, Here we go (points at PBS site). Hot mantle rises up from the mantle and triggered the weak parts of the crust within the interior of the continental plate. Yeah, that’s good. Alice winces. Frank: You don’t think so? Alice: Yeah, hold on a second, hot mantle material. Frank: Let’s go back to the first page. That was the best one.

The man in this pair takes the lead. He produces most of the claims himself, but also involves his partner with questions, and gives her an opportunity to make responses. He also makes a number of evaluative comments about the sources. Here is an example of the interaction of the final pair, Pair 4:

Rick: Alright, the first one is like because of plate tectonics. Lucy: Want me to go back to that one? Rick: Yes, I know how to describe it so basically the continental disturbance and the ocean floor are elements that cause earthquakes and stuff like that and lava and pressure build up. Lucy: So it causes disturbance and that makes the magma? Rick: Yeah, that thing right there leads into this, about subduction.
This pair has the most even production of writing content and claims. They question each other, they evaluate the sources, and they strategically plan their reading to a greater extent than the other pairs.

**TABLE 12.3**

Number of Discourse Moves by Each Pair and Partner

<table>
<thead>
<tr>
<th>Task</th>
<th>Summary</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Pair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total utterances</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Total questions</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total polite expressions</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Evaluative remarks</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Metacognitive remarks</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Planning of reading</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Planning of task</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Negotiation of content</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Claims</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Evidence from text</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Evidence from knowledge</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Reading Time on Reliable and Unreliable Sources

Another interesting variable to examine is the time spent reading the different sources across pairs. The sources were divided into reliable (credible authorities with accounts of volcanic activity that converged on plate tectonic theory) and unreliable (commercial or personal sites) sources. Pairs in the argument condition did tend to spend more time on reliable sources ($M = 1417$ seconds, $SD = 30$) than pairs in the summary condition ($M = 982$ seconds, $SD = 356$). However, on average, pairs in both conditions spent a similar amount of time on unreliable sources (argument $M = 587.5$, summary $M = 597$). These reading time results suggest that the argument task may be prompting students to spend a larger amount of time on relevant information than they would otherwise, but it is not generally improving the ability to discriminate and disregard irrelevant accounts.

Absence of Behaviors

Finally, it is interesting to note what behaviors were not observed in our pairs. In these protocols there was an absence of criticism of others’ statements, no direct conflict, no direct comparison of theories and evidence, and few attempts to develop coherent and sufficient models of volcanic activity. It is perhaps this lack of deeper evaluation of evidence and theories that relates to the overassimilation and inclusion of causes that we saw in all four essays. Even our best collaborating pair (Pair 4) considered unreliable information (the sources on astrology, the tides, and oil drilling) and no student directly questioned the reliability of these sources, or how these explanations fit with the other theories they were reading. Finally, there were few references to prior knowledge, which a student might have used as a source of evidence. Almost all of the evidence that was considered was from the text that the students had in front of them.

GENERAL DISCUSSION

This line of research has been concerned with investigating the kinds of learning contexts that may promote the best understanding when learning from technology, such as web pages. At least at first blush, the results of this study suggest that argumentation tasks may be a promising design element for collaborative learning environments. A related area that we can draw our attention to is how we may use the technology more interactively to further support collaborative learning. In the introduction, we suggested that the critical features of successful collaborative learning contexts could be that they maximize participation, interaction, and investment of all indi-
viduals; provide a grounds for evaluation that makes criticism less personal; give students the tools by which to accept or reject ideas as they attempt to supporting a thesis or interpretation; and help students to focus on relevant and reliable information. The exploratory study included here seems to support some of these notions, and specifically that argumentation tasks may in particular prompt more interaction and co-construction of understanding. Interestingly, in these dyads, an argumentation task did not seem to prompt more extensive evaluation of the reliability of sources, or more critical evaluation of claims or evidence. In turn, these may be prime candidates for direct instruction or prompting within a technology-supported collaborative learning environment. Technology can contribute perhaps by playing an interventionist role, such as by promoting recognition of conflict, or by forcing justification or explicit discrimination between theories.

These results are consistent with some findings in the social psychology literature that suggest that triads and not dyads may be the optimal size group for collaborative learning and problem solving (Steiner, 1972). Although the small size of a dyad is beneficial for maintaining motivation and participation, pairs tend to avoid conflict. For many reasons, differences in two-person groups tend to be glossed over rather than addressed head on. Triads, on the other hand, afford conflict by their number, and any difference in opinion leads to majority and minority positions, which in turn requires each member to evaluate those positions. Hence, one potentially advantageous role that a computer might play is as a third member of a triad. This may be better than a human triad, as the computer can assume the role of critic, push minority positions, and prompt evaluation and justification. At the same time, the computer partner will not monopolize the conversation, or read aloud while the partners are trying to think, or perhaps lessen the motivation of the two human participants. In preserving the inherent advantages of human and peer interaction, a human pair interacting with a computer may have advantages over models of a computer as learning partner where a single student is paired with a computer interface.

In this study, the students did not seem to take advantage of their partner as a catalyst for active evaluation of the available evidence. We do not know if there was anything about the dyad learning context that might have made either member of the pair less likely to achieve an understanding. The possibility remains that interacting in groups can make learning more difficult due to interruptions in each member’s thought process. To the extent that this is true, then there are other ways in which technology may improve collaborative learning, as there are several areas in which virtual collaboration may supersede face-to-face settings. Within an educational context, Linn and Hsi (2000) showed that locating student discussions in online chat groups allows all members to contribute more. Similarly, there is a parallel electronic brainstorming literature that has demonstrated some advantages for virtual groups over face-to-face collaboration (Gallup, Cooper, Grise, & Bastianutti, 1994; Valacich, Dennis, & Connolly, 1994). For example, Paulus and Yang (2000) found that putting brainstorming
sessions online minimizes the interruptive effects of others comments, but still allows for members to prime ideas and facilitate the cognition of others. These electronic methods of supporting participation so that each student has the ability to pursue his or her line of thought, while still contributing to a discussion, seem promising. On the other hand, attempts at structuring online dialogue have yet to result in learning gains (Hron, Hesse, Cress, & Giovis, 2000). Further, the effects of virtual collaboration are also likely to be highly specific to context. In direct opposition to Paulus and Yang's (2002) results, Straus and McGrath (1994) found that virtual groups can be less productive than face-to-face groups. As a result, although the use of technology to support collaborative learning is certainly a promising area of study, there is much work to be done before we can definitively assert which contexts may lead to effective group performance, what contexts should be supported with technology, and what allows group members to learn more successfully than they would alone.

ACKNOWLEDGMENTS

This research was supported by grants to the first author from the Office of Naval Research and the National Science Foundation. We thank Ivan Ash and Cara Jolly for their assistance with this project.

REFERENCES


