Appear in Proceedings of Computer Supported Collaborative Learning, 2007

Thinking Hard Together: the Long and Short of Collaborative Idea Generation in Scientific Inquiry

Hao-Chuan Wang, Carolyn P. Rosé, Yue Cui, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh PA 15213
Email: haochuan@cs.cmu.edu, cprose@cs.cmu.edu, ycui@cs.cmu.edu
Chun-Yen Chang, National Taiwan Normal University, 88, Sec. 4, Ting-Chou Rd., Taipei Taiwan 116
Email: changcy@ntnu.edu.tw
Chun-Chieh Huang, Tsai-Yen Li, National Chengchi University, 64, Sec. 2, Zhi-Nan Rd., Taipei Taiwan 116
Email: g9415@cs.nccu.edu.tw, li@nccu.edu.tw

Abstract: Idea generation is a cognitive process that plays a central role in inquiry learning tasks. This paper presents results from a controlled experiment in which we investigate the affect on productivity and learning from doing idea generation tasks individually versus in pairs, with versus without automatic support from a virtual brainstorming agent called VIBRANT. Our finding is that individuals brainstorming with VIBRANT produced more ideas than individuals who brainstormed with a human peer. However, an additional finding is that while brainstorming in pairs lead to short term process losses in terms of idea generation, with a corresponding reduction in learning in terms of pre to post test gains, it produced a productivity gain for a subsequent distinct individual inquiry task. Furthermore, automatically generated feedback from VIBRANT improved learning during idea generation but did not mitigate the process losses that were associated with reduced learning in the pairs conditions.

Introduction

Inquiry as an approach to learning typically consists of such activities as exploring the targeted phenomena, formulating and asking questions, making discoveries, achieving deeper understanding, and fulfilling intellectual curiosity. Virtually every inquiry activity begins with "asking questions" after which students may be requested to move on to "finding answers" or "testing the solutions", and subsequently, "asking better questions". *Idea generation* is of central importance in this process.

We are conducting our investigation in connection with the Debris Flow Hazard task (DFH), which is an example of an inquiry problem used by science educators as an assessment of creative problem solving ability (Chang & Weng 2002). The DFH task is defined by the following two idea generation prompts: *"What are the possible factors that might cause a debris-flow hazard to happen?"*, and subsequently, *"How could we prevent it from happening?"* Notice that the goal for students here is not to select and then apply a known procedure for solving a well defined problem. In contrast, the purpose here is to let students work first to define the problem and then creatively formulate the candidate problem solving steps/options. Beyond offering students the opportunity to generate possible solutions to problems, these tasks offer students the opportunity to weigh and balance trade-offs between alternative solutions since there is no single correct solution to the problem.

Based on cognitive theories of associative memory, idea generation can be viewed as the process of building on the retrieval of information encoded in a stimulated portion of a semantic network stored in one's long-term memory (Brown & Paulus, 2002; Dugosh et al., 2000; Nijstad & Stroebe, 2006). When students have access to domain facts either through their own memory or provided externally through access to learning resources, students may engage in a constructive process to bridge instances of domain facts on the way towards generating ideas (Brown & Paulus, 2002). For example, students may have access to the following two domain facts: (1) Debris flow refers to the mass movement of rocks and sedimentary materials in a fluid like manner. And, (2) There are many typhoons, or hurricanes, in Taiwan in the summer time. Students may then make the following two bridging inferences: (1) Heavy rain implies the presence of a massive amount of water. And, (2) The presence of a massive amount of water may lead to erosion or the movement of rocks in a fluid like manner. They may then generate the following idea: "Typhoons may be a factor leading to the occurrence of a debris flow hazard." As students are generating these bridging inferences, they are elaborating their mental representation of the basic facts they are building on. This process of building bridging inferences and subsequently elaborating mental representations is similar in many ways to the process of self explanation (Chi et al., 1994). In the learning sciences, self-explanation has been shown to be an effective learning process. Thus, through this constructive idea generation process, we expect to find a relationship between idea generation and learning much like the one that has been shown in many contexts between self-explanation and learning, and in fact we did find such a relationship, which we discuss below.

While idea generation in groups is purported to be more effective than idea generation for individuals, it is a well known problem that when groups engage in idea generation together, a phenomenon referred to as *process loss* occurs. In particular, it has been repeatedly demonstrated that a group that is interacting while doing idea generation together may not always perform better than a collection of non-interacting individuals whose contributions are simply pooled afterwards (i.e., nominal groups), both in terms of the quantity and quality of unique ideas, and in fact may sometimes perform significantly worse (Hill, 1982; Diehl & Stroebe, 1987; Nijstad & Stroebe, 2006). Often inquiry learning tasks such as the DFH task are done collaboratively in the classroom. To the extent that learning in inquiry tasks may come from the constructive process of generating ideas, we expect that factors that negatively affect idea generation productivity, such as the presence of evaluative statements (Dugosh et al., 2000) or exposure to instances of ideas that are close to the current idea generation focus (Nijstad & Stroebe, 2006), will also have a negative effect on learning from inquiry tasks where idea generation is involved. As we discuss below, we did find such a pattern in our data, which argues that the phenomenon of process losses in idea generation is a problem that should be taken seriously by learning scientists. Nevertheless, learning in idea generation tasks may arise from multiple different mechanisms, not only from the idea generation process per se. For example, while evaluative statements may inhibit productivity in idea generation, they count as a form of transactivity in collaborative discourse, which shows that group members are attending to one another's contributions and making explicit links between their contributions and those that came before. Supporting such behavior has been shown in other work to support learning (Weinberger et al., 2005).

While much research has been done separately on learning from inquiry tasks in the learning sciences community and the problem of process losses in connection with group idea generation in the social psychology of group work, in this paper we bring these two lines of research together to explore a particular question: How do the process losses that are a well known problem for group idea generation impact learning from inquiry tasks? And furthermore, how can we support learning by mitigating these process losses? Or do we gain more in terms of learning by enhancing other processes at work that may lead to learning even if they inhibit idea generation? In the remainder of this paper we formally explore the connection between learning and idea generation in inquiry tasks through an experimental study. While the results show that even with automatic idea generation support, we still see evidence of process losses connected with a loss in learning, we do see a positive effect on learning of the automatic support mechanism we introduce. Furthermore, we find a positive impact of collaborative idea generation on preparation for a subsequent idea generation task.

Hypotheses and Model

The hypotheses underlying our investigation grow out of the social psychology literature on creativity and group brainstorming as well as the cognitive science literature on associative memory and collaborative learning. The model presented in Figure 1 depicts the hypothesized causal links between interventions (i.e., whether students worked with feedback from the VIBRANT agent or not, and whether they worked in pairs or individually), mediating variables (i.e., cognitive stimulation and social interaction), and dependent measures important in inquiry learning tasks (i.e., productivity in idea production and learning). In the figure, a "+" symbol denotes a positive influence imposed by the node at the initial end of the arrow on the node at final the end of the arrow, while a "++" symbol represents a qualitatively stronger positive influence, and a "-" symbol denotes a negative influence. Circled numbers are included to enhance clarity. Link (a) represents the positive effect of priming stimuli on associative memory activation (Brown & Paulus, 2002; Dugosh et al., 2000). Link (b) denotes the potential learning benefit of knowledge construction (analogous to the process of self-explanation) triggered by the idea generation process (Chi et al., 1994). Link (c) is an inhibitory influence on idea generation, possibly due to a diversion from pure idea generation by evaluative conversation or elaboration, or exposure to instances of ideas too similar to the current focus of idea generation (Nijstad & Stroebe, 2006). Link (d) represents a predicted positive influence of interaction on learning, consistent with reported advantages of collaborative learning (e.g., Weinberger et al., 2005).

From this model, we derive four specific hypotheses that we explore subsequently in an experimental study: (1) Working in pairs will have a differential effect on productivity and learning such that students in the pairs condition will be less productive in their brainstorming but may still learn more. (2) Working with the support of the VIBRANT agent, which provides stimulation in the form of reference to general categories of ideas, will be more effective for stimulating idea production than working with a human peer to the extent that human peers primarily provide concrete instances of ideas rather than general categories of ideas (Nijstad & Stroebe, 2006). (3) Feedback during problem solving supports learning, thus we hypothesize that students working with the VIBRANT agent will evidence more domain learning than students in the no support conditions (Bangert-Drowns et al., 1991). (4) Transactive social interaction supports the acquisition of multi-perspective knowledge (Weinberger, 2003), thus we hypothesize that students in the pairs condition will be more effective at a subsequent idea generation task that builds on ideas discussed in the first brainstorming task.

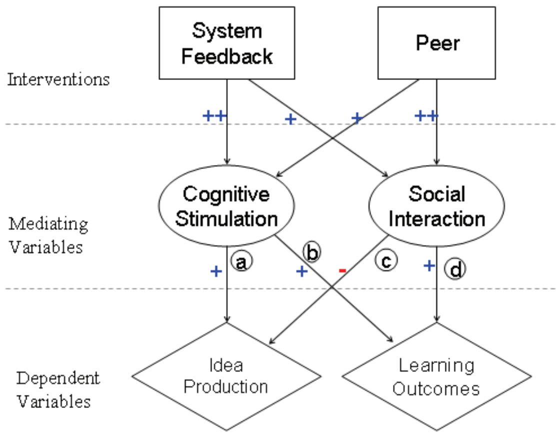

Figure 1. An influence diagram depicting hypothesized causal connections between interventions, constructs and outcome measures.

Method

Experimental Design

In order to test our hypotheses, we conducted an experiment in which students participated in a brainstorming task in an educational context. The Debris Flow Hazard (DFH) task, which is the brainstorming task we selected, has been designed by science educators to engage students in scientific inquiry in the area of Earth sciences (Chang & Tsai, 2005). The learning objective of this task is to make concepts related to geology, agriculture, and urban development concrete for students as they grapple with the manner in which these very different types of factors interact in real world scenarios. However, it is more similar in its cognitive demands to other idea generation tasks used in studies of group dynamics than typical collaborative learning tasks such as mathematics problem solving or collaborative writing. Thus, the specific properties of this task make it particularly appropriate for beginning to explore the separate and joint effects of cognitive and social factors on the productivity and pedagogical value of brainstorming activities. We manipulated whether brainstorming took place as an individual or pair activity and whether feedback was offered or not, both as between subjects factors. Thus, the experiment was a 2 (individual brainstorming vs. pair brainstorming) X 2 (no system support vs. system support) factorial design resulting in four experimental conditions, which are referred to in the remainder of the paper as IN (Individual-No support), IS (Individual-System supported), PN (Pair-No support), and PS (Pair-System supported).

Experimental Infrastructure

In order to implement the four conditions in a way that maintains maximal consistency across conditions, we built our experimental infrastructure on top of a well known instant messaging (IM) service over the Internet, Microsoft Network's MSN messenger (msn.com). Due to the popularity of this IM service with the target user population, using an MSN-based client also lessens potential concerns of software difficulty or novelty effects.

We adapted an existing brainstorming feedback agent called VIBRANT (Wang et al, 2006) to provide prompts in response to conversational behavior in the two system supported conditions. In order to be adapted to a specific task, VIBRANT must be provided with an idea hierarchy at multiple levels of abstraction. In our domain idea hierarchy, the top node representing the entire DFH task is first broken down into 5 general topic areas including geology (e.g., shale rock area), agriculture (e.g., having shallow-rooted economic plants which cannot solidify the soil mass as much as original forests), influences caused by other natural phenomena (e.g., typhoon and rainstorm which break the hydraulic balance), urban development (e.g., building houses at a potential dangerous slope), and social factors (e.g., improper environmental policy). Each subtopic is further broken down into specific idea nodes. A total of 19 specific idea nodes are included. Feedback messages are attached to the nodes of the idea hierarchy both at the general topic level and the specific idea level. Similarly, at the specific idea level, prototype expressions of the related idea collected in previous studies involving the DFH task are attached to idea nodes. In this way, student conversational contributions can be matched to nodes in the hierarchy by matching the text of their contribution to the associated prototype texts using a simple semantic similarity measure.

The feedback provided by VIBRANT consists of two parts. The initial portion, which we refer to as the *comment*, acknowledges the idea that matched, and how it fits or doesn't fit into the hierarchy. The second portion, which we refer to as the *tutorial*, offers a hint for thinking about a new contribution. Feedback messages are constructed by concatenating a selected comment with a selected tutorial. For example, if the student has contributed the idea "deforestation", the system will acknowledge this with the following comment, "Good, you seem familiar with the effects of excessive urban development." A next focus for brainstorming, which coherently follows from this would be more discussion related to urban development, for example "Can you think of a farming practice motivated by economic concerns that may increase the risk of a debris flow hazard?" VIBRANT never offers students specific ideas. Instead, the hints offered by VIBRANT are more similar to the "category label" stimuli (such as "improve parking" for the task of "how can your university be improved?") demonstrated to enhance idea production in previous studies of group and individual brainstorming (Dugosh et al., 2000; Nijstad & Stroebe, 2006). VIBRANT's built in strategy for selecting a next focus was designed to balance breadth and depth of brainstorming across the idea hierarchy while maintaining the coherence of the conversation. This design is motivated by prior findings that brainstorming is more efficient when successive ideas are clustered so that semantically related ideas are contributed in close proximity, and transitions between general idea categories are relatively rare (Nijstad & Stroebe, 2006).

For the IS condition, VIBRANT offered feedback in response to each contribution of the student. For the PS condition, in order to give students time to react to each other's contributions before viewing automatically generated feedback, the system collected and evaluated the two students' contributions during a fixed period of time, and then gave feedback based on the accumulated text. This adjustment of the parameter, length of time for collecting dialogues, may be viewed as adjusting how interruptive the computer agent is. In this study, the parameter was set to 30 seconds, which was observed during a pilot experiment to allow students enough time to interact with one another. No feedback from the system was offered to students in the two no support conditions. Thus, in contrast to the two support conditions just described, for the IN condition, a simple computer agent did nothing but simply recorded students' contributions. Students were simply instructed to use the IM program as a text input buffer. A similar simple agent was used in the PN condition where pairs of students brainstormed together on the IM platform but received no system support.

Participants

The study was conducted in a computer classroom of a public high school located in central Taiwan. Four sessions were scheduled in the same day, two in the morning and two in the afternoon. In each session, the computer classroom accommodated at most 16 students. Every student worked at a computer assigned to him or her. Participating students were allowed to choose the session they attended, and were randomly assigned to experimental conditions within that session. For experimental conditions PN and PS, students were paired into dyads randomly. The version of the MSN based software used as our experimental infrastructure was configured so that in the pair conditions, the other student assigned to the same dyad a student appeared in the "buddy list" of that student. Additionally, in the support conditions, the computer agent that provides feedback also appeared in the "buddy list". Thus, when the students would launch the application during the experimental manipulation, their MSN based client would be configured to support a conversation between all of the relevant parties. Altogether, there were 7 students in the IN condition, 7 students in IS, 14 students in PN (i.e., 7 pairs), and 14 students in PS (i.e., 7 pairs). During the study, all students were blind to the experimental design, and unaware of the existence of other conditions.

Experimental Procedure

The experimental procedure can be divided into five phases, namely (1) background readings, (2) pretest, (3) brainstorming 1, (4) brainstorming 2, and (5) the post test. The experimental manipulation took place during phase (3), which is the first brainstorming phase. The purpose of the second brainstorming phase is to test whether the experimental manipulation from phase 3 has an effect on brainstorming behavior that can be detected within a new brainstorming task. While prior work has evaluated the effect of collaborative idea generation on a subsequent individual idea generation stage where the idea generation task was the same, to the best of our knowledge this is the first evaluation in an experimental study of the effect of collaborative idea generation on a subsequent different idea generation task. We strictly controlled for time in all phases.

Phase 1. Background Reading (10 minutes)

During phase 1, the students read a 3 page packet of background reading materials on the climate, geology, and development of Taiwan as well as some information about natural disasters but no specific information about debris flow hazards. This packet was compiled by domain experts working in a science education center at National Taiwan Normal University. The readings were designed to offer students a wide range of background material related to the topics contained within the idea hierarchy discussed above, however it did not contain the direct answers to any questions on the test nor did it directly express the ideas students were required to contribute in the brainstorming task. The reading material itself does not explicitly introduce the factors underlying DFH occurrences. The purpose of the reading materials was to prepare student for the brainstorming task. Students in all conditions were instructed to read the material for 10 minutes, and to learn as much as possible from the material. The readings were given to students prior to the pretest so that any learning measured by pre to post-test gains can be attributed to the brainstorming task and not to the readings alone. At the end of the 10 minutes, students were asked to turn the reading materials over and not look at them. Lab attendants ensured that students followed the instructions.

Phase 2. Pre-brainstorming Test (15 minutes)

In phase 2, students took an on-line pretest assessing their conceptual knowledge and reasoning about debris flow hazards.

Phase 3. Brainstorming Activity 1 (30 minutes)

After the pretest, the students participated in the first brainstorming phase, which is where the experimental manipulation took place. Students were instructed to launch the MSN program and to start working on the DFH brainstorming task. Specific instructions for the task appeared as the first prompt in the MSN messenger window. Students were given a scenario about a specific debris flow hazard and then asked to generate as many thoughts as possible in answer to the question, "*what are the possible factors that may cause a debris flow hazard to happen?*" During this activity, students were invited to use the reading materials from Phase 1 as a resource. The duration of the brainstorming session was limited to 30 minutes.

Phase 4. Brainstorming Activity 2 (10 minutes)

Upon the completion of the brainstorming task, students regardless of experimental condition were then instructed to do individual brainstorming on a second brainstorming task. In this idea generation task, students were requested to offer preventive solutions for DFH. The prompt for this solution-finding brainstorming activity was "*what facilities or solutions may prevent a debris flow hazard from happening?*" No system support, reading material or peer interaction was provided when doing this transfer task. The purpose of this task was to assess whether the impact of the experimental manipulation had a lasting effect beyond the duration of the manipulation.

Phase 5. Post-brainstorming Test (15 minutes)

Finally, students took an on-line post-test identical to the one used as a pretest again in order to assess the influence of the experimental manipulation on learning outcomes. The time allowed for doing the test is also the same to the pretest phase (for 15 minutes).

Measurement

Three outcome measures were used and analyzed in this study, including pre to post-test learning gains, productivity during the initial idea generation task during the experimental manipulation, and productivity in a subsequent idea generation task.

We use a 26-item domain test for assessing students' concept comprehension on the DFH topic. The test itself can conceptually be further decomposed into two parts, factual knowledge recall questions (11 items) and more reasoning-oriented questions (15 items). The test served as the indicator of students' learning status at the pre-test and post-test phases. The test was designed by science education researchers for high school students and has been used in previous science education studies (Chang et al., in press). The validity and reliability of this instrument were discussed and established in prior studies (Chang et al., in press).

We examined the productivity of idea generation in the first brainstorming activity by using two counting methods: number of unique ideas contributed by each student and total number of unique ideas produced by groups as a whole. The first task performance measure was the number of unique ideas generated by each individual student. Students' brainstorming contributions are coded and classified to one of the 19 ideas modeled in the aforementioned idea hierarchy. Duplicate ideas are ignored in this analysis. For students who brainstormed with peers in the PN and PS conditions, we only counted an idea as a unique idea that student contributed if that student was the one who mentioned it first. The second performance measure we looked at is group-based idea production which is standard to studies of group idea generation in the literature (Diehl & Storebe, 1987). Unlike the first measure, here we looked at the output of brainstorming groups as a whole rather than that of individual group members. For students in the two individual sessions without a human peer (i.e., IS and IN), for a fair comparison, we formed 7 "nominal" dyads for IN and 7 for IS in a posthoc manner by randomly selecting two individuals from the same experimental condition (either IS or IN). Ideas generated by group members of nominal dyads were pooled for the comparison with real groups.

For the transfer task, students were evaluated based on the number of unique ideas they were able to contribute during the allotted time. Only ideas that matched a list of valid ideas collected during previous studies using this task counted in the unique idea count.

Results

Four hypotheses proposed at the beginning were evaluated and examined in the following analyses based on measures of conceptual learning, performance in the main brainstorming activity, and performance of the subsequent idea generation task.

Data Coding

Logs of all IM behavior in all conditions were saved for analysis. Altogether we collected 28 logs, 7 in each condition. Note that in the pairs condition, there is only one log per pair rather. To derive appropriate quantitative measures of idea generation for analyses, including task performance (number of unique ideas in the main idea generation task) and transfer performance (number of unique ideas the solution-finding transfer task), data collected in the main brainstorming phase (i.e., phase 3) and the transfer task phase (i.e., phase 4) have been coded.

For the main idea generation task, student IM conversation logs were first segmented into idea units, since during IM conversations, students may contribute more than one idea per turn. The inter-rater reliability between two independent coders over 10% of the data for sentence segmentation was satisfactory (Kappa= .7). Each unit contribution was then classified into one of the 19 domain concepts in the aforementioned idea hierarchy. If there was no feasible label for a particular contribution, the label of "other" was given. The inter-rater reliability for the concept coding over 10% of the data was also sufficiently high (Kappa=.84). As described previously, the number of unique ideas generated by each individual and the transformed efficiency measure (number of unique ideas/number of total unit sentences) were both computed from this coding. For the second brainstorming task, students' responses to were coded according to a coding scheme developed by domain experts based on prior studies. The categories in that coding scheme represent 15 valuable ideas. The inter-rater reliability of this coding of two independent coders over 10% of the data was Kappa=.74, which is satisfactory.

In order to gain insights about the social process of idea generation, the conversation logs were also coded on the social dimension of knowledge communication. A coding scheme consisting of 11 classes was developed. There are six valuable classes in the coding scheme which include *elaboration* (i.e., idea justification or explanation), *comment*, *positive evaluation*, *negative evaluation*, *question* (i.e., seeking for explanations), and *suggestion*. These valuable classes are considered as indicators that group members actively engaged in exchanging on-task information, arguing with each other, and co-constructing knowledge collaboratively. We also identified other four social codes which can be roughly characterized as off-task social interactions, including *encouragement*, *greeting*, *acknowledgement,* and *meaningless utterance*. Finally, for unit contributions which are solely idea instances carrying no social mode, we coded them as *idea*. The inter-rater reliability was acceptable (Kappa=.75).

Hypothesis 1: Differential Effect of Social Interaction on Idea Generation and Conceptual Learning

The prediction of hypothesis 1 consists of two parts. The first part predicts the presence of productivity loss in brainstorming groups, and therefore students who worked in pairs should be less productive in idea generation. The second part predicts that social interaction may promote conceptual learning. Thus, it is expected that students in the Pairs conditions would gain more knowledge as measured by the domain test.

Hypothesis 1-a: Productivity Loss in Pairs

Because production blocking is a well-known explanation for productivity loss in group brainstorming (Diehl & Stroebe, 1987), which may effect both idea production and learning, we began our analysis by investigating whether there was evidence of significant production blocking (i.e., having fewer chances to contribute ideas due to turn-taking) in our data either from the presence of a peer or from the involvement of a computer agent in the conversation, but we did not find evidence of this. We first computed an ANOVA with the two independent factors from our experimental manipulation as the independent variables and total number of student contributions (i.e., regardless whether they contain ideas or not) as the dependent variable. The ANOVA did not show a significant effect of the Individual/Pair factor, and in fact the trend was for students in the pairs condition to make more conversational contributions than students in the individual conditions. Similarly, not only did we not find evidence of production blocking due to the presence of a computer agent in the conversation, we found a marginal main effect in favor of System Support associated with the System-support/No-support factor, $F(1, 38)=3.62$, $p<.1$, with a medium effect size $f=.26$ (Cohen's $f=.25$-$.40$, or equivalently, Cohen's $d= .50$-$.80$) (pp. 286-287, Cohen, 1988) (System Support- Mean: 30.43, S.D.: 15.01; No Support- Mean: 19.95, S.D.: 10.24), demonstrating that the trend was in the opposite direction of what would be predicted if there were production blocking based on this very rough measure of production blocking. Thus, we do not find evidence that the presence of either a human or computer partner for brainstorming reduces the opportunity for students to contribute to the conversation.

Although there was no evidence of production blocking, in the analyses we still find evidence of productivity loss from the Pairs conditions when we use unique ideas matching one of the 19 ideas selected by science educators for this task. The primary ANOVA model was set up by using the first performance measure that we have mentioned in the following way:

(A-1) D.V.: Number of Unique Ideas by Each Student, I.V.: Individual/Pairs, System-Support/No-Support

A significant main effect for Individual/Pair in favor of *individual brainstorming* was found, $F(1,38)=70.94$, $p<.001$, Cohen's $f=1.37$ is very large (Individual- Mean: 9.57, S.D.: 1.91; Pair- Mean: 4.61, S.D.: 1.73). With respect to the other independent factor, the presence of adaptive feedback generated by VIBRANT seemed to have a trend benefiting the number of unique ideas but did not result in significant difference. No interaction effect was found. We also looked at the group-based production performance by using the second productivity measure, in which we formed nominal groups for experimental conditions IN and IS, and then pooled ideas generated by nominal group members statistically. By using the group-based measure, a significant main effect on the comparison of nominal groups versus interacting groups (i.e., real groups, PN and PS conditions) was found, $F(1, 24)= 20.7$, $p<.001$, $f= .93$, which is still large (Nominal Pair- Mean: 12.36, S.D.:1.55; Real Pair- Mean: 9.21, S.D.:2.12).

Hypothesis 1-b: Learning Outcomes in Pairs

In connection with conceptual learning, we first evaluated the general learning outcomes in terms of concept comprehension by computing a repeated measures ANOVA with time point (pre versus post test) as an independent factor. From this analysis we determined that there was a main effect of time point with no two-way or three-way interactions with our experimental manipulation. $F(1,76)= 9.35$, $p < .005$, Cohen's $f = .35$, which is a medium to large effect size (Pretest- Mean: 7.41, S.D.: 1.32; Posttest- Mean: 8.14, S.D.: 1.35). Thus, we conclude that students across conditions learned significantly from pretest to posttest in the brainstorming activity.

Then we examined the effect of our experimental manipulation on the magnitude of learning. We hypothesized that because of the benefits of collaborative learning interactions, we would see a learning benefit for collaborative idea generation even in the face of process losses with respect to productivity oriented outcome measures. We did find evidence of an increase in the number of instances of the types of conversational contributions we expected to be associated with learning in the pairs condition based on our analysis of the corpus, especially in the condition where the pair of students interacted with the VIBRANT agent. In particular, by counting the number of valuable social contributions that we have annotated (i.e., the six task-related social codes.), through an ANOVA analysis, it is determined that students who worked in pairs produced more valuable social interactions than students who worked individually, $F(1, 38)=5.1$, $p<.05$, a medium to large effect size f=.37 (Pair-Mean: 8.96, S.D.: 9.20; Individual- Mean: 3.36, S.D.: 4.45). There was no main effect of system support and no interaction effect. Table 1 further shows the average numbers of social contribution per category that occurred per session in each experimental condition. Note that number of ideas in this table refers to any idea contribution, whether it was unique or not, and whether it matched one of the 19 pre-specified ideas or not. From the top of the table, it appears that for the two Individual conditions, students narrowly focused on the core idea generation task. Very few extended explanations or other social interactions were ever uttered. Even in the PN condition the number of valuable social codes (e.g.., elaboration, comment, positive/negative evaluation, question and suggestion) is still low, and *not* significantly different from their occurrence in the Individual conditions. From the comparison between PN and PS, it is noteworthy that though our system feedback did not explicitly prompt students to engage in social interactions, such as elaborations, comment and evaluations etc., the current feedback seemed to indirectly trigger more intense social interaction within the pairs condition. Students in the PS condition have significantly more social contributions than those in other conditions especially in connection with several particular social codes, such as elaboration, comment and positive/negative evaluation that are associated with transactive collaborative discourse (Weinberger et al., 2005). The result can be viewed as the evidence that interaction with a collaboration support agent such as VIBRANT can lead to an increase in transactivity (Weinberger, 2003) in the ensuing collaborative discourse, however the impact was not great enough to lead to increased learning when comparing IS with PS.

Table 1. The mean and S.D. of various social contributions per student per session in each experimental condition. Superscripts indicate statistically different levels of occurrence.

	IN	IS	PN	PS
Idea	20.71 (4.96) ^{A}	21.00 (5.35) ^{A}	11.00 (4.10) ^{B}	17.36 (9.88) $^{A\,B}$
Elaboration*	1.14 (2.19) $^{A\,B}$	0 ^{B}	.64 (1.34) $^{A\,B}$	1.36 (1.28) ^{A}
Comment*	2.43 (3.60) $^{A\,B}$	1.29 (2.21) ^{B}	2.86 (3.68) ^{B}	6.71 (6.50) ^{A}
Positive Evaluation	0 ^{B}	1.86 (2.04) $^{A\,B}$	.57 (1.16) ^{B}	2.43 (2.53) ^{A}
Negative Evaluation	0 ^{B}	0 ^{B}	0 ^{B}	.57 (.76) ^{A}
Question	0	0	1.07 (3.25)	.79 (1.37)
Suggestion	0	0	.29 (.61)	.64 (1.45)
Acknowledgement	0	0	.71(1.64)	1.07 (1.54)
Greeting	0	0	0	.57 (1.16)
Encouragement	0	0	.14 (.53)	.07 (.27)
Meaningless	0 ^{B}	0 ^{B}	.5 (1.09) $^{A\,B}$	2.00 (2.48) ^{A}

*Student's t test, others: Tukey ; Values not labeled with the same letter are significantly different

We then evaluated whether the increase in occurrence of valuable social contributions in the Pairs condition were associated with increased learning as measured by pre to post test gains. We did this using an ANCOVA analysis configured as below:

(A-2) D.V.: Total posttest score, I.V.: Individual/Pairs, System-Support/No-Support, Covariate: Pretest score

There was a significant main effect of System Support, $F(1, 38)=4.57$, $p<.05$, Cohen's f = .35, which is a medium to large effect. Students in the system-supported conditions achieved significantly higher *adjusted posttest scores* (System Support- Mean: 8.61, Std. Err: .20; No Support- Mean: 8.02, Std. Err: .21). A significant main effect was also found on the factor Individual versus Pair $F(1,38)=12.17$, p=<.01, effect size Cohen's f=.84, which is a large effect. Students who brainstormed individually without a peer learned significantly better (Individual- Mean: 8.85, Std. Err: .25; No Support- Mean: 7.78, Std. Err: .17). No statistical interaction effect was found between the two independent variables. The ranking of adjusted posttest scores for the four experimental conditions is: IS (Mean: 9.05, Std. Err: .34) > IN (Mean: 8.66, Std. Err: .37) > PS (Mean: 8.12, Std. Err: .24) > PN (Mean: 7.43, Std. Err: .24). Students learned most in the IS condition, in which VIBRANT adaptive feedback was available, while no peer was present. However, only the difference between the two extreme conditions (IS and PN) is significant based on a Bonferroni post-hoc analysis. Students who brainstormed with the VIBRANT agent learned significantly more than students who brainstormed with a peer and no system support.

Though the phenomenon of productivity loss was observed consistent with hypothesis 1, in connection with conceptual learning, a pattern in opposition to hypothesis 1's prediction was obtained. Students did not learn conceptual knowledge better due to social interaction, and in fact, students in the Pairs conditions learned significantly less. A further exploration on the relation between idea production and learning outcomes revealed a correlation between the two measures. By classifying students into two groups according to a median split of their numbers of unique ideas generated, and using the domain pre-test as the covariate, it was found students with higher numbers of unique ideas scored significantly higher on the domain post-test, $F(1, 39)=9.03$, $p<.01$, a large effect size Cohen's $f=.48$. Students with more ideas scored better in the domain test (More productive brainstormer- Mean: 8.66, Std. Err: .23; Less productive brainstormer- Mean: 7.75, Std. Err: .20).

Hypothesis 2: Effect of Categorical Cognitive Stimuli

Hypothesis 2 predicts that students who brainstormed alone with the VIBRANT agent (i.e., the IS condition) would be more productive than students working purely in pairs (i.e., the PN condition). The comparison of IS and PN can be viewed as comparing the how different types of stimuli affect the cognitive process of idea generation. Both conditions had only one source of stimulation, either carefully designed stimuli from a computer agent or naturally occurring stimuli from a human peer. A Bonferroni post-hoc test on a previously introduced ANOVA model (A-1) showed that there was a significant difference between IS and PN on their productivity during brainstorming 1 (IS- Mean: 10.14, S.D.: 1.95; PN- Mean: 4.50, S.D: 1.91). While interaction with a human peer leads to a significant decrease in idea production, we do not see this effect resulting from interaction with the VIBRANT agent, and in fact the trend is in the opposite direction. This finding supports hypothesis 2.

Hypothesis 3: Adaptive feedback as Learning Support

Hypothesis 3 predicts that students would learn domain concepts from the VIBRANT's adaptive feedback intended as cognitive stimuli for brainstorming. From the ANCOVA model (A-2) introduced previously, it was determined that students learned significantly better when adaptive feedback was available. No interaction with other variables was found. Hypothesis 3 was supported by this result.

Hypothesis 4: Effect of Social Interaction on Subsequent Idea Generation

We hypothesized that students who worked in pairs in the main brainstorming task would be more effective at a subsequent related but different idea generation task. We first examined the relation between performance in the subsequent idea generation task and other measures. The number of unique ideas was used as the measure for this transfer task. No significant relation was found between measures of the transfer task and the main task. Nevertheless, by categorizing students into two groups, High/Low reasoning ability in the domain, according to a median split on their performance on the reasoning-oriented part of the domain test, students with high reasoning ability in the domain were determined to be more capable in the second idea generation task, $F(1, 40)=4.28$, $p<.05$, a medium to large effect size Cohen's $f=.33$ (High reasoning- Mean: 5.75, S.D.: 1.89; Low reasoning: Mean: 4.77, S.D.: 1.11).

A two way ANOVA was conducted by using the number of unique solutions as the dependent variable, experimental manipulations as independent variables, and the aforementioned label on High/Low domain reasoning ability was added into the ANOVA model to account for variance related to that factor. A significant main effect was found for the Individual/Pair factor, $F(1, 37)=7.67$, $p<.01$, a large effect size $f=.46$. The result was in favor of *working in pairs* (Pair- Mean: 5.54, S.D.: 1.58; Individual- Mean: 4.64, S.D.: 1.50). Also, a significant interaction effect was detected between our two experimentally manipulated factors, $F(1, 37)=5.57$, $p<.05$, $f=.39$, which is close to a large effect size. PS was found to be the best condition in the transfer task (Mean: 5.79, SD: 1.72), while IS was the worst (Mean: 4.00, SD: 1.41). A post-hoc pair-wise Bonferroni analysis showed that PS and PN both had significantly better performance than IS in the transfer task.

Hypothesis 4 was therefore supported. Students in the Pairs conditions preformed better in a subsequent idea generation session, in which a related but different task became the target and no external support was available.

Conclusions and Current Directions

We have presented the results of an experimental study investigating both the long term and short term effects of brainstorming in pairs versus brainstorming individually in an inquiry learning context. Our finding is that brainstorming tasks can be beneficial for student learning. Furthermore, because of a significant correlation between brainstorming productivity and pre to post test learning in our data, the results support the view that learning from brainstorming comes from the constructive process of idea generation. Beyond that, the condition favored by the results depends upon what outcome measure is valued above the others. For example, students in the pairs condition were less productive and learned less during the initial brainstorming task. On the other hand, the students who brainstormed in pairs during the first session performed better on the second brainstorming task. Furthermore, although brainstorming support had a positive effect on learning both in the individual and pairs conditions, it did not have a significant positive effect on productivity during the initial brainstorming session. Nevertheless, since high reasoning ability students performed better on the second task, we see the learning gains especially related to the reasoning portion of the pre/post test that resulted from the feedback as providing a potential lasting positive effect on future brainstorming. If the relationship between idea generation and learning can be verified to be a causal one, a brainstorming agent that better supports productivity may also better support learning. Since students in the pair conditions were observed to have many repetitions and paraphrases of the same ideas, one potential future agent design might be one that encourages partners to explore different parts of the idea space to avoid producing redundant ideas, and potentially to avoid process losses due to cognitive interference (c.f., Nijstad & Stroebe, 2006).

This work was funded in part by the National Science Council (NSC), Republic of China grant number NSC 94-2511-S-003-028 and National Science Foundation grant number SBE0354420.

References

Bangert-Drowns, R., Kulik, C., Kulik, J., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events, *Review of Educational Research*, 61(2), pp 213-238.

Brown, V. R., & Paulus, P. B. (2002). Making group brainstorming more effective: recommendations from an associative memory perspective. *Current Directions in Psychological Science*. 11(6), 208-212.

Chang, C., & Weng, Y. (2002). An exploratory study on students' problem-solving ability in earth science. *International Journal of Science Education,* 24(5). 441-451.

Chang C., & Tsai, C. (2005). The interplay between different forms of CAI and students' preferences of learning environment in the secondary science class. *Science Education*, 89, 707-724.

Chang, C., Barufaldi, J. P., Lin, M., Chen, Y. (in press). Assessing tenth-grade students' problem solving ability in the area of Earth sciences. *Computers in Human Behavior*.

Chi, M. H., De Leeuw, N., Chiu, M., Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-488.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Ed.*. NJ: Lawrence Erlbaum Associates.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology*. 53(3), 497-509.

Dugosh, K. L., Paulus, P. B., Roland, E. J., & Yang, H. (2000). Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology.* 79(5), 722-735.

Hill, G. W. (1982). Group versus individual performance: are *N*+1 heads better than one? *Psychological Bulletin*, 91(3), 517-539.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: a cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3), 186-213.

Wang, H., Li, T., Huang, C., Chang, C., Rosé, C. P. (2006). VIBRANT: A brainstorming agent for computer supported creative problem solving, *Proceedings of the Intelligent Tutoring Systems Conference (ITS 2006)*.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33* (1), 1-30.