JrnlID 11412_ArtID 9124_Proof# 1 - 12/06/2011

Computer-Supported Collaborative Learning
DOI 10.1007/s11412-011-9124-x

Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry

Hao-Chuan Wang · Carolyn P. Rosé · Chun-Yen Chang

Received: 27 August 2010 / Accepted: 7 June 2011

Abstract This paper seeks to contribute new insight to the process of learning during idea generation (i.e., brainstorming) by proposing and evaluating two alternative operationalizations for learning, which we refer to as connection-based learning and multi-perspective learning, during a carefully designed idea-generation task in the earth-sciences domain. Specifically, this paper presents two controlled experiments. In the first study we manipulate two independent factors, first whether students work individually or in pairs, and second whether students work with the VIBRANT agent or not. The second study includes one additional hybrid agent condition motivated by results from the first study as well as other enhancements to the VIBRANT agent's discussion-analysis technology. Our finding is that while brainstorming in pairs leads to short-term process losses in terms of idea-generation productivity, with a corresponding reduction in connection-based learning, it produces a gain in multi-perspective learning. Furthermore, automatically generated feedback from VIBRANT improves connection-based learning. In the second study, support from an enhanced version of VIBRANT showed evidence of mitigating the process losses that were associated with reduced learning in the pairs condition of the first study.

Keywords Collaborative idea generation · Collaborative process analysis · Dynamic collaboration support

H.-C. Wang
Cornell University, 301 College Ave., Ithica, NY 14850, USA
e-mail: haochuan@cs.cornell.edu

C. P. Rosé (✉)
Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
e-mail: cprose@cs.cmu.edu

C.-Y. Chang
National Taiwan Normal University, 88, Sec. 4, Ting-Chou Rd., Taipei, Taiwan 116
e-mail: changcy@ntnu.edu.tw

Introduction Q1

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. *Idea generation* (i.e., brainstorming) is of central importance in this process, and frequently these idea-generation tasks are done collaboratively. Despite the overwhelming evidence of process losses (i.e., when individuals function less productively in a group than individually) during group brainstorming (Connolly Q2 1993; Diehl and Stroebe 1987; Kraut 2003), the reality of modern life is that realistic idea-generation tasks in the workplace often must involve more than one individual, often from diverse backgrounds, such as in multi-disciplinary and frequently multi-national design teams. An important question in collaborative group work is how to support productive idea generation even in the face of tendencies towards process losses that plague realistic working environments.

The process of learning during collaboration and the process of collaboratively producing high volume output or a high quality product are separate processes that may occur at the same time but may be at odds with one another. Emphasizing one of these goals, such as short-term productivity, may lead to a loss with respect to the other goal. For example, under realistic working conditions in order to speed up short-term progress towards a solution, groups may fall into dysfunctional communication patterns such as quick consensus-building behavior (Weinberger and Fischer 2006) or resort to divide-and-conquer problem-solving approaches where team members work in relative isolation on the part of the process they already know. As a result, team members do not have the opportunity to exchange ideas and gain valuable multi-perspective knowledge (Weinberger et al. 2005) or learn new skills. Perhaps more importantly, this dysfunctional group communication can lead to design flaws (Dutoit 1996), which tend to be discovered late in a development process when they are expensive to fix (NIST 2002).

Much of the social psychology literature on group brainstorming emphasizes short term productivity. However, an alternative perspective would be to view successful groups as ones that strike an appropriate balance between high productivity in the short term and learning in preparation for future work, which may lead to better long-term performance. For example, evidence from empirical studies suggests that brainstorming in a group might lead to a large "productivity spike" (Brown and Paulus 2002) in a subsequent individual brainstorming session on the same topic. In this paper, we explore the topic of collaborative idea generation both from a short-term and a long -term perspective. From the short-term perspective success in a brainstorming task is measured in terms of the number of unique, high quality ideas that are produced. But from a longer-term perspective, success may be evaluated in terms of learning about the problem that may occur during brainstorming or in terms of cognitive preparation for productivity during a subsequent distinct brainstorming task. A long-term design goal is to support collaborative idea generation in such as way as to maximize long-term benefits while minimizing short-term process losses. One way of doing this is using conversational agents that participate in group discussions along with the students.

We present the results from two behavioral studies in which we evaluate variations on a design of virtual brainstorming support involving a conversational agent called VIBRANT, whose design is motivated by prior work in the area of collaborative idea generation (Nijstad and Stroebe 2006). In the first study we manipulate two independent factors, first whether students work individually or in pairs, and second whether students work with the VIBRANT agent or not. The second study includes one additional hybrid agent condition motivated by results from the first study as well as other enhancements to the VIBRANT agent's discussion analysis technology. 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? One positive contribution of this work is a demonstrated connection between idea generation and learning in a carefully constructed task. On the negative side, the results from the studies show that even with automatic idea generation support, we still see evidence of process losses connected with a loss in learning, although 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. An important take away message is that the literature on group process losses is important for the CSCL community to consider in the design of collaborative tasks and that both the positive and negative effects of group interaction need to be carefully balanced and managed.

Theoretical foundation and hypotheses

Brainstorming is an activity that is frequently listed among those activities that CSCL environments are meant to provide a venue for. However, while it has been purported to be a learning task, we know of few empirical studies that offer an evidence base for an understanding of how brainstorming leads to learning. In this paper, we propose two operationalizations of learning during brainstorming, the first we refer to as *connection-based learning* and the second we refer to as *multi-perspective learning*.

Brainstorming may be modeled from a cognitive perspective as a two stage process in which a search is initiated in response to a challenge during the first stage, and then an idea is constructed by means of inferences building upon prior knowledge stimulated through this search process (Nijstad and Stroebe 2006). Psychological studies in spreading activation have demonstrated that it is more efficient to retrieve a concept from memory when *relevant* information is offered as a *prime* or cue (Anderson 2005; Raaijmakers and Shiffrin 1981). Based on the cognitive theory of associative memory (Anderson 2005; Brown and Paulus 2002; Dugosh et al. 2000), idea generation can be viewed as the process building on the retrieval of information encoded in the stimulated portions of a semantic network stored in one's long-term memory. New ideas are generated when bridging inferences are made between these now salient pieces of knowledge. In addition to producing an idea, the bridging inference leads to an enrichment in the cognitive representation of the salient knowledge in connection with the newly generated idea. This enrichment is what we refer to as connection-based learning. Stimulation that broadens the search for ideas in brainstorming (Dugosh et al. 2000) may increase the likelihood of cognitive conflict. Cognitive conflict is the mental state in which learners become conscious of gaps in their understanding, which increases their receptivity to cognitive restructuring and learning (Piaget 1985).

On the other hand, it is possible that the learning during brainstorming is not due to the act of generating ideas itself, but due to the exposure to other perspectives that enrich the representation of the domain that participating students have. We refer to this second operationalization of learning as *multi-perspective learning*. Within the collaborative learning community there is much support for the benefits of learning collaboratively over learning alone (Sharan 1980; etc.). Just a few specific examples include investigations in mathematics problem solving (Gweon et al. 2006) and in conceptual learning for electronics (Gokhale 1995). Gweon et al (2006) also demonstrate that an appropriate intervention for drawing out and elaborating conversational interactions can further enhance the learning benefits students receive through peer interaction. A series of studies in the computer- supported collaborative learning field demonstrate the pedagogical value of social interaction by showing that an intervention that intensifies argumentative knowledge construction in support of consensus building in the context of group work enhances the development of multi-perspective knowledge, where students learn to view a problem from multiple angles (Weinberger et al. 2005; Weinberger and Fischer 2006). Weinberger and Fischer measure evidence in multi-perspective learning by investigating the extent to which an analysis contains evidence of looking at a situation from more than one point of view. And they report an association between productive argumentative knowledge construction and multi-perspective learning. Thus, to the extent that this type of interaction occurs during group brainstorming, we may find evidence of this multi-perspective learning as a result of the interaction that occurs during group brainstorming. This learning may be impervious to the process losses that are well known within the social psychology literature.

Even with evidence that brainstorming is a learning task, it would be reasonable to question whether it is really a collaborative task if group brainstorming is prone to process losses that make group brainstorming less efficient than individual brainstorming. Where brainstorming is less efficient, we would expect less connection- based learning. Prior work in the area of the social psychology of cooperative work leads us to expect such an effect. The task of idea generation in brainstorming groups has been extensively studied through controlled experiments (Diehl and Stroebe 1987) and simulation studies (Nijstad and Stroebe 2006). This empirical work has repeatedly revealed phenomena related to process losses, in which a group with idea sharing 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 (Hill 1982; Diehl and Stroebe 1987). A wide range of explanations for process losses have been proposed and tested empirically, including social pressure (e.g., evaluation apprehension), social loafing (e.g., "free riding"), and production blocking resulting from turn taking conventions (Connolly 1993; Diehl and Stroebe 1987; Kraut 2003). Q4 Idea failure associated with these process losses may be related to cognitive interference that occurs at several stages (Wang and Rosé 2007). (Nijstad and Stroebe 2006), such as when one fails to retrieve an image (a subset of semantically interrelated concepts) from memory, retrieves an image but fails to generate an idea from it, repeatedly activates the same images, or generates ideas that have already been mentioned. When brainstorming in groups, overhearing a peer's ideas may serve as cognitive stimulation for memory retrieval and idea generation when those ideas are significantly different from what an individual was capable of generating alone, while on the other hand, this external stimulation may become a source of cognitive interference leading to process losses. This may occur in the case that one student generates an idea based on knowledge stimulated in his mind. His partner hears this idea and then generates a similar idea. However, upon hearing this similar idea, the first student's brain activates very similar knowledge to what was already activated. Because the same domain knowledge continues to be activated, this facilitates the generation of similar ideas to those mentioned recently. In cases such as these, brainstorming individuals may believe they have exhausted the number of ideas they are able to generate (Nijstad and Stroebe 2006).

Our investigation is guided by two hypotheses, which are not mutually exclusive:

- Hypothesis 1 is that connection-based learning occurs during brainstorming as a result of the act of generating ideas during brainstorming itself. If this hypothesis finds support in our data, then we expect to see learning negatively affected by the kinds of process losses reported in the group work literature. We thus expect that dynamic support that directly addresses the process losses would increase learning by means of increasing brainstorming productivity.
- Hypothesis 2 is that multi-perspective learning occurs during brainstorming as a result of the social interaction between students, during which students are exposed to alternative perspectives, which may have the effect of enriching their conceptual understanding. If this hypothesis finds support in our data, we would not necessarily expect learning to be negatively affected by process losses. In this case, the ideal support for group brainstorming would seek to enrich the social interaction between students rather than seeking to circumvent the process losses per se.

If we do not see evidence of learning from brainstorming connected with either of these operationalizations, then we fail to find support for brainstorming as a learning task. Before we test these hypotheses, we begin by further explicating these two complementary operationalizations of learning and how they are afforded by a carefully designed brainstorming task.

Method

We present two studies using the same experimental paradigm in which we evaluate the connection-based learning hypothesis (Hypothesis 1) and the multi-perspective learning hypothesis (Hypothesis 2). In both cases, one factor manipulates whether brainstorming occurs as an individual or pair activity, and are thus able to isolate the effect of process losses. In accordance with prior work, we expect to see process losses in the paired condition. If hypothesis 1 is correct, we expect to observe connection-based learning during the brainstorming task, but that students learn less in the pairs condition and that this learning loss effect is mediated by productivity in the brainstorming task. If hypothesis 2 is correct, we expect to see an increase in multi-perspective learning in the pairs condition in comparison to the individual condition. A second factor in both studies is a feedback manipulation in which support is given in the form of “category labels”, which were designed to address process losses. If hypothesis 1 is correct, and if the feedback is effective at mitigating process losses, it is also expected to mitigate the corresponding learning loss. What is different between the two studies is the mechanism used to determine the selection and timing of the feedback with respect to the task context. The result of this manipulation differs between studies, offering insight into the importance of context appropriateness of brainstorming support as well as insight into what technological approaches achieve the most effective performance.

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 posttest. The

JrnlID 11412_ArtID 9124_Proof# 1 - 12/06/2011

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 that takes place in phase 3 has a lasting effect on brainstorming behavior beyond the duration of the manipulation 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 a behavioral study of the effect of collaborative idea generation on a subsequent different idea generation task. We strictly controlled for time in all phases. Here we describe the whole procedure in detail.

Phase 1. Background reading (10 min)

Students in all conditions were instructed to read the 3-page supplemental reading material designed to give them background on the geology of Taiwan for 10 min, 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 posttest gains can be attributed to the brainstorming task and not to the readings alone. At the end of the 10 min, 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 min)

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

Phase 3. Brainstorming activity 1 (25 min)

The brainstorming task that provides the context for both of the investigations of learning during collaborative idea generation that we report on is the Debris Flow Hazard (DFH) task. This task has been designed by science educators to engage students in scientific inquiry and creative problem solving in the area of Earth sciences (Chang and Weng 2002). 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 chat client program and to start working on the DFH brainstorming task, described in detail below. Specific instructions for the task appeared as the first prompt in the chat 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 25 min.

Phase 4. Brainstorming activity 2 (10 min)

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 min)

Finally, students took an on-line posttest 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 min).

Task design

The learning objective of the Debris Flow Hazard (DFH) 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 exploring the separate and joint effects of cognitive and social factors on the productivity and pedagogical value of brainstorming activities.

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?*" The duration of the brainstorming session was limited to 25 min. Regardless of condition, students participated in the brainstorming session through typed chat in using a chat client in the style of the Microsoft Network Messenger (MSN messenger). Similar to MSN messenger, turn taking was not enforced so that students were constantly free to enter ideas even when their partner was also typing. Upon the completion of the brainstorming task, students regardless of experimental condition were then instructed to do individual brainstorming on a second brainstorming task, this time on paper. In this second 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?*"

As part of the task as mentioned in connection with Phase 1 above, students are provided with 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 or how to prevent them. For example, in discussion of the climate of Taiwan, the reading states the following:

> Taiwan is located where west Pacific typhoons (hurricanes) frequently pass, and thus typhoons often visit Taiwan during the summer season. A typhoon is a tropical cyclone developing from the disturbance of the tropical atmosphere. … One characteristic of a typhoon is its huge amount of rainfall. Rainfall accompanying typhoons can account for more than half of the annual precipitation in Taiwan, and thus often causes great damage.

The packet aims to provide basic information about the natural environment of Taiwan, and the information may support reasoning on many problems, including debris flow hazards. However, the design of the reading materials is not specific to the issue of debris flow or its related hazards. Note that a debris flow hazard is not a natural phenomenon. It is the situation where a naturally occurring phenomenon (i.e., heavy rains) becomes dangerous to people and/or property because of careless decisions made by people, such as choosing to live in an area where mudslides are likely to occur or removing vegetation that would make the environment more stable against mudslides. These issues were not explicitly covered in the packet. This packet was compiled by domain experts working in the science education center at National Taiwan Normal University. The purpose of the reading materials was to prepare students for the brainstorming task but not to give them specific answers.

The packet is essential for making this idea generation task be one in which conceptual learning can take place. When external information sources are provided to students in support of their brainstorming but not directly contributing answers to the brainstorming, then the students must engage in a constructive process much like self-explanation (Chi et al. 1994) or reflection based learning (Gustafson and Bennett 1999) in order to use this information for idea generation. Self-explanation is a learning process during which people make inferences to connect new information to prior knowledge, and thereby integrate multiple sources of information. From a scientific viewpoint, one of the best substantiated educational findings in cognitive science research is the educational benefit of explanation, and in particular, the self-explanation effect (Chi et al. 1994; Renkl 2002). Self-explanation benefits learners by revealing knowledge gaps, abstracting problem specific knowledge into schemas that can be applied to other relevant cases, and elaborating the representation of knowledge in the learner's mind so that it can be more easily retrieved (VanLehn and Jones 1993). To the extent that idea generation prompts cognitive processes similar to self-explanation, idea generation may have pedagogical value. For example, a student may have access to the following two domain facts discussed in the reading materials: (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. That student 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. That student may then generate the following idea: "Typhoons may be a factor leading to the occurrence of a debris flow hazard." This enriches the student's mental representation of the connection between environmental factors and debris flow hazards as well as contributing towards success at the idea generation task.

Measuring learning and productivity

There are many skills and concepts that could be learned during brainstorming, but we focused on two in particular. One is conceptual learning about the domain as measured by a standardized test. We refer to this as connection-based learning because the learning activity is designed to provide opportunities for this learning through the connections students make between details from the readings and their real-world knowledge during brainstorming. The other learning measure we refer to is multi-perspective learning, which we operationalize as preparation for a subsequent idea generation task, as measured by productivity within that second task. Here we do not mean that we are teaching creativity or ability to generate ideas on an arbitrary idea generation task. Instead, we focus on a specific type of preparation where the second task builds directly on the first task and success is determined based on the extent to which students are able to generate solutions from multiple perspectives. Whereas in the first task, students worked on problematization, in the second task, they generated ideas for avoiding the identified problematic situations. In this section we address both how we operationalize learning and how we operationalize brainstorming productivity.

Measuring productivity

The DFH task has been piloted in classroom studies, and the ideas generated by students in those prior trials were recorded and analyzed by a panel of science educators in order to identify the reoccurring ideas that they considered valid ideas. While no constraints are placed on the range of ideas students are able to generate during the task, in practice it rarely happens that students generate a valid idea that is not one that has been seen frequently in prior studies. Altogether, 19 valid re-occurring ideas were identified for the first task and 15 for the second task, each of which were organized into an idea hierarchy that captures the relationships between ideas. During task 1, the ideas contributed by the students were recorded in logs saved by the chat client. For task 2, the ideas generated by the students were written on paper. These records produced during the brainstorming tasks were then used for analysis.

The first task performance measure was the number of unique ideas generated by each individual student. Students' brainstorming contributions during the chat were coded and classified according to one of the 19 ideas modeled in the aforementioned idea hierarchy. Duplicate ideas are ignored in this analysis. For students who brainstorm with peers, we only count an idea as a unique idea that student contributed if that student is the one who mentions it first. The second performance measure was group-based idea production, which is standard in studies of group idea generation in the literature (Diehl and Stroebe 1987). When we compare individuals and pairs, it would not be fair to directly compare the number of ideas produced by two students together with what can be produced by one student working alone. Thus, we adopt the standard approach used in the group work literature. For students who do idea generation alone, we form "nominal" dyads by randomly selecting a partner from the pool of students who worked individually. We then pool the ideas generated by both students in this "nominal dyad" and count the unique ideas within that pool.

Measuring connection-based learning

As an assessment of understanding of domain concepts, which we used both as a pretest and a posttest, we adopted a standardized assessment developed by science educators (Chang et al. 2007), which was a 26-item multiple choice test designed for assessing students' concept comprehension on the Debris Flow Hazard topic but did not directly questions reasons for the occurrence of debris flow hazards. The test itself can conceptually be further decomposed into two parts, factual knowledge recall questions (11 items) and reasoning-oriented questions (15 items). The test was designed for high school students and has been used in previous science education studies. The validity and reliability of this instrument were discussed and established in prior studies (Chang et al. 2007). One sample item of test reads as the following:

> Debris flows often occur while encountering typhoons. What is the most appropriate description of the relation between debris flow and typhoons?

(a) Seawater encroachment raised by typhoons may then erode the shoulder of mountain slopes.
(b) The heavy wind may intensify the weathering of the rock, and destroy the rock formations.
(c) The intense rainfall accompanying typhoons may then carry a lot of soil and rocks, which then slide down the slope.

(d) The wind of typhoons is so strong that it causes soil and rocks to fall down.

Recall that the readings they were given for the students to use as a resource offering them a wide range of background material related to relevant topics for their task, 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. It did discuss aspects related to the geology of Taiwan, such as its size, topology, and climate.

Because we used this test both as a pretest, which occurred before the brainstorming activity, and as a posttest, after the brainstorming activity, we cannot eliminate the possibility that some of the information on the test itself may have primed students for the brainstorming task. However, because the pretest occurred strictly before any experimental manipulation, we can be certain that any priming effect that it did have did not differ between conditions and thus does not interfere with our ability to assess the effect of our experimental manipulation.

Measuring multi-perspective learning

Multi-perspective learning is acquiring the ability to view a problem from multiple perspectives in order to enhance the ability to generate multiple alternative solutions to a problem. Since the learning task is focused on defining the problem of debris flow hazards, an appropriate test of multi-perspective learning would be a second brainstorming task about solving the identified problems. Thus, we operationalize multi-perspective learning as preparation for a subsequent idea generation task in which students are asked to generate solutions to the problem of debris flow hazards, and we measure this learning in terms of productivity within that second task. As with our measure of productivity on the first brainstorming task, only ideas that matched a list of valid ideas collected during previous studies using this task counted in the unique idea count.

Verbal protocol analysis

Logs of all IM behavior in all conditions from both studies were saved for analysis with respect to idea generation. Note that in the pairs condition, there is only one log per pair rather than one log per student. 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 idea generation task), data collected in the main brainstorming phase (phase 3) and the transfer task phase (phase 4) were coded and inventoried.

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 idea categories 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). Similarly, for the second brainstorming task, students' responses were coded according to a coding scheme developed by domain experts based on prior studies. The inter-rater reliability of this coding of two independent coders over 10% of the data was Kappa=.74, which is satisfactory.

After we coded the data from the first study by hand, we experimented with a tool for automatic analysis of conversational data called TagHelper tools[1] to check whether we could improve the performance of idea identification. Our finding was that we were able to achieve an agreement of automatically predicted topic labels with human assigned topic levels at a high reliability, specifically .7 Kappa, which was considerably higher than the performance of the original VIBRANT system's analysis component, which achieved a Kappa of only .5 in comparison with human coding. Thus, we replaced VIBRANT's internal topic identification software with this TagHelper model for the second study. After the second study, we analyzed the new data both fully automatically using the TagHelper model that was trained on the previous study's hand annotated data as well as by hand and evaluated the agreement. Again we achieved a Kappa of .73, demonstrating that automatic analysis of this type of conversational data is feasible with high reliability and generalizes to different pools of students than those it was trained on.

Technical infrastructure: The vibrant agent

In order to maintain maximal consistency across experimental conditions in the two studies we report in this paper, we built our experimental infrastructure displayed in Fig. 1 on top of a well known instant messaging (IM) service over the Internet, MSN.

We adapted a brainstorming feedback agent developed in our prior intelligent tutoring work called VIBRANT (Wang et al. 2006; Kumar et al. 2007) to provide prompts in response to conversational behavior in the experimental conditions that include system support. The same chat client was used in both studies reported in this paper for all students, regardless of condition. The only thing that changed was who was participating in the chat, i.e., whether there was one or two students, and whether there was also a computer agent participating. Figure 1 displays a setup where a pair of students and a computer agent are working together.

Because the range of valid ideas that students generate during this task can be easily enumerated, we are able to organize the 19 ideas associated with the first task into a domain hierarchy. 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 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., poor environmental policy). Each subtopic is further broken down into specific idea nodes. A total of 19 specific idea nodes are included.

The VIBRANT feedback generation approach is similar in spirit to that adopted in prior work in tutorial dialogue, namely the Geometry Explanation Tutor (Popescu et al. 2003) and Auto-Tutor (Graesser et al. 2001) projects. However, our approach differs from this prior work in several important respects. First, similar to Popescu et al. (2003), we attach feedback messages to nodes in our hierarchy so that we can use a match between a student contribution and a node in the hierarchy as a basis for selecting a feedback message. However, in contrast to Popescu et al. (2003), we do not utilize a deep, symbolic approach to language understanding. Instead, we employ two alternative light weight approaches. In

[1] For the automatic analysis we used a publicly available verbal analysis toolset called TagHelper tools, available at http://www.cs.cmu.edu/~cprose/TagHelper.html.

JrnlID 11412_ArtID 9124_Proof# 1 - 12/06/2011

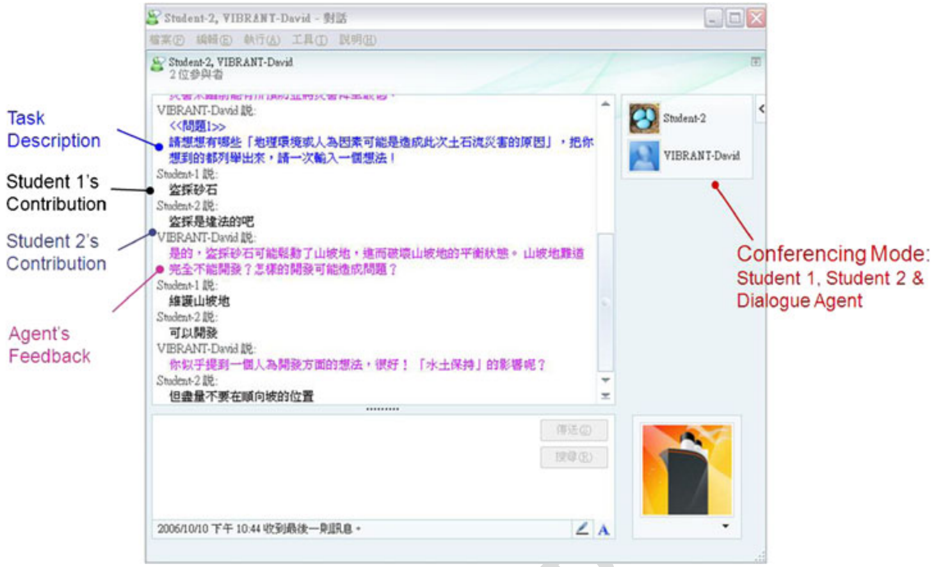

Fig. 1 The MSN like chat client used in the study. Note that students interacted with the agent in Chinese. An example translated into English is displayed below in Fig. 2

study 1, we associate a number of prototype texts to each leaf node in the hierarchy so that we can determine which node matches best based on a shallow semantic similarity measure. This approach is much lighter weight than the earlier approach adopted by Popescue and colleagues in that neither require heavy knowledge representation or inferencing technology.

Similar to Graesser et al. (2001) we make use of a finite state machine to determine how to use the hierarchy to select feedback. However, in contrast to the Auto-Tutor approach, our strategy is motivated more by general principles of dialogue coherence rather than a specific knowledge elicitation strategy designed to elicit a specific idea from a student with progressively more pointed hints. 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 and Stroebe 2006).

VIBRANT's feedback design is based on the idea of "category labels" (Dugosh et al. 2000; Nijstad and Stroebe 2006) which have been investigated in the context of earlier work on group brainstorming. For example, in the context of a brainstorming task where participants are generating design ideas for improving a university campus, an example "category label" would be "improve parking". This represents a type of idea rather than a specific concrete idea, such as "improve parking by converting the football field into a parking lot." Results from evaluations of category label stimuli delivered by human experimenters show that it is effective for increasing cognitive stimulation and idea generation productivity. Our experimental infrastructure illustrates how a group brainstorming environment can be built that automatically provides stimulation in the form of category labels and goes beyond that prior work by enabling that feedback to be generated automatically in a context sensitive manner.

The *feedback* produced by the system consists of two parts. The first part is referred to as a *Comment*, which is meant to offer feedback specifically directed at the details of what students have contributed. The comment text associated with the most closely matching idea node is selected, unless the goodness of the match is rated as low, in which case a more abstract node that subsumes that idea node is selected instead. The second portion of the feedback is referred to as a *Tutorial*, the purpose of which is to direct the student to a new idea, preferably which coherently follows from the current idea if such a subsequent focus exists in the hierarchy and has not yet been covered by the student. 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?” In this way students have the opportunity to learn how to evaluate their ideas, and they are encouraged to continue contributing additional ideas. Two separate Finite State Machines are designed for the selection of each of the two feedback portions, which then appear as two separate sentences communicated to the student(s). In the current design, a *Comment* is selected based on a function only of the current student contribution, while a *Tutorial* is selected based on a function of the student contribution and the context of students’ previous responses within the same session. The technical details of this approach to feedback generation are discussed in a previous publication (Kumar et al. 2007).

In Fig. 2 we see an example with a pair of students working together with the VIBRANT agent. Here, one student has mentioned the problem that a dangerous situation with respect to debris flow hazards is created when people remove sand and stones from construction sites, since this makes the area less stable. The agent acknowledges this idea in the comment, mentioning material that was included in the reading materials students had access to in support of the brainstorming. A tutorial is then offered, directing the students on how to proceed with the brainstorming. In this example, we see the first student offering an example and then the second student offering a synergistic example. This pattern was much more frequent in the supported pairs condition than the unsupported pairs condition. Figure 3 shows another example of a pair of students interacting with the agent. In this example, a student first mentions that inherent environmental conditions, like the angle of a mountain slope, could be a factor for the occurrence of debris flows. The agent recognizes the statement as an idea related to inappropriate construction on slopes, and prompts the students to think about a related topic of mining on mountain slopes. Students were then motivated to explore the role of people in debris flow hazards, and proposed an idea that growing shallow-rooted betel trees may be an explanatory factor. Interestingly, although the agent did not perfectly recognize and follow the notion of “steep angle slope” suggested by

Fig. 2 An example with a pair of students working with the VIBRANT agent

Student 1	People stole sand and stones to use for construction.
Agent	Yes, steeling sand and stones may destroy the balance and thus make mountain areas unstable. Thinking about development of mountain areas, can you think of a kind of development that may cause a problem?
Student 2	Development of mountain areas often causes problems.
Student 1	It is okay to develop, but there must be some constraints.

Fig. 3 Another example of a pair of students interacting with the agent

Student 1	It may be caused by factors of the geographical environment, such as the slope angle of a hill.
Agent	Yes, one should not build buildings at inappropriate locations. Is the happening of a debris flow relevant to the mining of sand and stones?
Student 2	I felt that factors related to people play a bigger role.
Student 1	Yes, people-related factors would be the most influential.
Student 2	Such as planting many betel palm trees on mountain slopes.

the student, the feedback provided by the agent helped the two students to identify a new direction of thinking and move beyond their original focus.

The agent may support individual idea generation in a similar way. Figure 4 shows how individual students may work with the agent. Here, a single student follows the agent's tutorial feedback closely to propose ideas. After the agent directs the student to think about various topics (e.g., mining sand and stones, problematic uses of slopes etc.), the student responds to the ideas with enriched reasoning along the directions suggested (e.g., mining without approval, inadvisable farming practices, etc.). In this case, there appears to be rich connection-based thinking, although with no social interaction and less exploration of diverse perspectives.

Study 1: Investigating the problem of process losses

Experimental manipulation

In order to test our hypotheses, we conducted an experiment in which students participated in a brainstorming task in an educational context. 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).

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

Fig. 4 An example of a single student interacting with the agent

Student	Rainfall adds weight to the soil and stones, and the gravity causes it to slide down.
Agent	Yes, a mountain slope can only bear limited amount of pressure exerted by objects and buildings upon it. Is the happening of a debris flow relevant to the mining of sand and stone?
Student	People mine sand and stones without approval.
Agent	Yes, stealing sand and stones without approval can loosen the mountain slope, and then destruct the status of balance. Is it the case that there cannot be any urban development on mountain slopes? What kind of development may cause a problem?
Student	It may be a problem to grow fruit trees and tea trees.

feedback based on the accumulated text. If that text contained more than one idea, then it is likely that more than one idea node would match. The idea node that got the best match score would be selected in that case. This adjustment of the parameter that controls the length of time for collecting text from students before intervening may be viewed as adjusting how interruptive the computer agent is. In this study, the parameter was set to 30 s, 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.

Subjects

Participating students were an approximately gender balanced group of 10th graders from a high school in Taipei, Taiwan. 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. 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. Students participated in the study as a learning activity connected with their regular instruction and thus did not receive financial compensation.

Results

Table 1 Q6

Hypothesis 1: Connection-based learning hypothesis

A summary of the results from study 1 can be found in Table 1. We will begin the discussion with results related to Hypothesis 1. The connection-based learning hypothesis states that students learn during brainstorming as an effect of the new connections they make between domain concepts as they generate ideas. If we observe a reduction in productivity as a result

Table 1 Summary of results from Study 1

	Individual—No Support	Individual—System Supported	Pair—No Support	Pair—System Supported
Pretest	7.9 (1.6)	7.3 (.9)	7.2 (1.4)	7.3 (1.5)
Posttest	8.8 (1.3)	9.0 (1.2)	7.3 (1.1)	8.0 (1.1)
Unique Ideas from Task1 (individual level analysis)	8.3 (2.3)	10.1 (2.0)	4.5 (1.9)	4.8 (1.5)
Unique Ideas from Task 2	5 (1.3)	4 (1.4)	5.3 (1.4)	5.6 (1.7)

of students working in pairs, we should also see a reduction in learning, and this reduction in learning should be at least partially explained by the reduction in productivity. Our data bears this out. Furthermore, we see an improvement in productivity as a result of brainstorming support, which is also associated with an increase in learning.

First, when we examine the productivity loss as a result of students working in pairs, we 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.

In terms of connection-based 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. Thus, we conclude that students across conditions learned significantly from pretest to posttest in the brainstorming activity.

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

Students who brainstormed individually without a peer learned significantly better. In order to determine whether this difference in learning was at least partially explained by the reduction in productivity, we explored the connection between idea production and learning outcomes revealed a correlation between the two measures that suggests that the process loss effect on idea generation productivity might explain the negative effect of group brainstorming on learning in comparison to individual brainstorming in this study. By classifying students into two groups according to a median split of their numbers of unique ideas generated, and using the domain pretest as the covariate, it was found that students with more unique ideas scored significantly higher on the domain posttest, F(1, 39)=9.04, $p<.01$, a large effect size Cohen's f=.48. Students with more ideas scored better in the domain test.

Thus, we find support for the connection-based learning hypothesis. The next question in connection with this hypothesis is whether feedback from the VIBRANT agent increased productivity and learning. At the individual level, we didn't find evidence that the support of the VIBRANT agent improved productivity. The presence of adaptive feedback generated by VIBRANT had a trend benefiting the number of unique ideas but did not result in a significant difference. No interaction effect was found. However, at the group level, we do see an effect. For this analysis we examined productivity by forming nominal groups for experimental conditions IN and IS, and then pooled ideas generated by nominal group member. 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 a large effect. Thus, there is some evidence from this analysis that support from VIBRANT improved productivity.

Based on this positive effect of VIBRANT feedback on productivity, hypothesis 1 predicts that we will see a positive effect on connection-based learning. From the ANCOVA model (A-2) introduced previously, it was determined that students learned significantly more when adaptive feedback was available. 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. No interaction with other variables was found. No significant 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.

Hypothesis 2: Multi-perspective learning hypothesis

The multi-perspective learning hypothesis states that a benefit of brainstorming in pairs is that students will gain more in terms of multi-perspective learning as a result of having been exposed to alternative perspectives. And we find support for this hypothesis in our data. As mentioned above, multi-perspective learning was measured in terms of productivity on the second brainstorming task.

Because we did not have a similar problem solving task to use as a pretest measure for this type of knowledge, we explored the relationships between success at the second task and other student specific information recorded prior to the experimental manipulation that we could use as a covariate in order to avoid treating all students as though they started with equal ability to perform this task. We found that score on the domain reasoning portion of the pretest was weakly correlated with productivity on the second brainstorming task and thus could serve such a purpose in our analysis. Specifically, when we categorize students into two groups, High/Low reasoning ability in the domain, according to a median split of their performance on the reasoning-oriented part of the domain pretest, 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 $f=.33$. Thus we controlled for this individual difference in the analysis of the effect of condition on productivity in the second brainstorming task.

An ANOVA was conducted using the number of unique solutions as the dependent variable, experimental manipulation as independent variables, and the aforementioned label on High/ Low domain reasoning ability as an additional factor. In this model, 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*. 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. Thus, while the brainstorming support improved brainstorming productivity in the first task both for individuals and pairs, it only contributed to multi-perspective learning in pairs, which makes sense if the multi-perspective learning comes from interaction with a peer rather than from productivity in the first brainstorming task.

Hypothesis 2 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.

Discussion

The results of the first study support both hypotheses about learning from idea generation. In particular, the results demonstrate that even in this simple brainstorming task, we see evidence of significant process losses that are associated with corresponding learning losses in terms of connection-based learning. And while support from the VIBRANT agent increases both productivity and learning, in this study we do not see that the effect is strong enough to ameliorate the process losses due to working in pairs. On the positive side, however, students who worked in pairs, especially with support, gained more in terms of multi-perspective learning.

The take away message from the study does not end there, however. With this result in hand, it is important to think again about the design of the learning task and how that enabled both connection-based learning and multi-perspective learning.

The conclusion with respect to multi-perspective learning was the most clear. We find support for this hypothesis in our data in the form of a main effect of the individual versus pairs manipulation. Thus, despite the process losses we observed, students were better prepared for a problem-solving task that built upon a problematization process that was done collaboratively. It's important to note that we consider this result to be very specific to the connection between problematization and problem solving. We do not consider that students who participate in a brainstorming task necessarily learn to be more creative on arbitrary idea generation tasks. Thus, we believe the application of this finding is within multi-step inquiry projects where students must first define a problem and then solve it. Our finding supports the idea that students will be able to be more creative about hypothesizing solutions to problems that they have a broader and deeper representation of from their problematization. This is consistent with the literature on productive failure where it has been found that groups that struggled with defining an ill-structured problem performed better on both near and far transfer assessment items than groups who worked on a more narrowly defined problem (Kapur and Kinzer 2009). A post-hoc analysis of the chat logs from the study showed that students in the pairs conditions with feedback stayed on topic longer than students in the other conditions, which gave them more opportunities to build on one another's ideas and think more about the implications of the ideas, as displayed in Figs. 2 and 3. In contrast, in the individual condition, the feedback increased the number of different ideas but not the amount of depth or breadth with which the ideas were considered, as displayed in Fig. 4.

With respect to connection-based learning, the hypothesis was that students would learn from idea generation in the Debris Flow Hazard Task because the act of generating ideas gives them the opportunity to form bridging inferences between their existing knowledge of debris flow hazards and the information provided to them in the supplementary readings. Assuming students engage in this process of constructing bridging inferences, we would expect the result to be an enriched representation of the domain, which we would then be able to measure using an assessment that depends upon that enrichment in order for students to score well. We saw that because the experimental setup allowed us to see that. Thus, while the experiment as designed allowed us to measure learning of this form, it is important to note that this result would not necessarily generalize to idea generation tasks that do not provide resources and opportunities for new bridging inferences to enrich conceptual knowledge, and would also not necessarily generalize to other knowledge associated with a domain that is not connected to the enrichment facilitated by the bridging inferences afforded by the task. Recent work on support for group idea generation in an engineering design setting showed that effective agent based support increased idea generation productivity, but not learning, and also did not find any correlation between productivity and learning (Kumar et al. 2011). We cannot therefore conclude that arbitrary idea generation tasks are learning tasks. What we can conclude is that carefully constructed idea generation tasks can facilitate very specific learning.

The connection between process losses and learning losses also give us pause as we consider whether idea generation should really be a collaborative task in a learning context. The support offered by VIBRANT was able to increase both idea generation and productivity, however, students who worked individually with the VIBRANT agent both produced more ideas and learned more than students who worked in pairs with the VIBRANT agent. Thus, while we observe a positive effect of this support, it was not observed in this study to fully mitigate the process losses.

Study 2: Follow-up study

Study 1 provided support for both hypotheses related to learning from idea generation. However, it did not leave us with a definitive idea of how best to design a collaborative idea generation task for learning. We were left with the question of how to strike the optimal balance of both forms of learning when multi-perspective learning comes from working on idea generation tasks in pairs, but that leads to process losses that impede connection-based learning. In a post-hoc analysis, we observed idea generation intensity as well as the magnitude of process losses from working in pairs to be highest in the first five minutes of idea generation. We also found that rather than improve idea generation productivity, the agent's intervention actually had the opposite effect. Thus we hypothesized that one possible solution would be to let students work alone without support for the first five minutes and then work in pairs supported by the agent for the remainder of the session. We also improved the accuracy of VIBRANT's idea detection capabilities after the end of the first study. The technique used in the first study was not always successful in identifying a satisfactory match to an idea node. In those cases, the fallback was to move up a level of abstraction in the idea hierarchy, and use the best matching more abstract category in order to partially compensate for the partial match. In study 2, we instead employed a machine learning approach inside the VIBRANT agent where we used labeled data from study 1 to train a model to do the assignment of novel input texts to nodes in the hierarchy. Note that apart from the idea matching, the VIBRANT agent was identical to that used in Study 1. We discussed this result in detail where the evaluation of corpus data collected during Study 1 was discussed above. The improvement in results was sufficient to remove the necessity for falling back to a more abstract partial match.

In the next section we describe a follow-up study in which we evaluate the effect of the enhanced VIBRANT agent as well as a new configuration in which students worked alone for the first five minutes, and then worked with a partner and the VIBRANT agent in the remainder of the session. We refer to this new condition as Dynamic (DYN), and motivate it from folk wisdom about brainstorming where some believe that brainstorming in groups is more effective when individuals take time to brainstorm alone first. Apart from the enhancement of VIBRANT and the additional condition, the study was identical to the first one, and thus, we spend less time discussing it. While the focus of the first study was to extend our understanding of learning during idea generation, the purpose of the second study was to help us better understand the practical side of supporting it with computer agent technology.

Experimental manipulation

As in the first study, 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). Additionally, we added a hybrid condition in which students brainstormed alone for 5 min without system support and then worked together for the remaining 25 min with system support (DYN). Earlier we mentioned that the VIBRANT agent was improved between the first and second studies in that the feedback for the second study was triggered by analysis of brainstorming activity by a text classification model trained by TagHelper tools (Rosé et al. 2008), which performed better than in the first study.

Subjects

Participating students were an approximately gender balanced group of 10th grade students from a high school in central Taiwan. They participated in the study as a summer school course activity and thus did not receive financial compensation for their participation. As in the first study, every student worked at a computer assigned to him or her. The study was conducted over two separate class periods of equal length of two different days. All five experimental conditions were equally represented in both sessions in order to control for possible systematic differences between the two groups of students. They were randomly assigned to experimental conditions within the session they participated in. For experimental conditions PN, PS and DYN, students were paired into dyads randomly. Altogether, the dataset we analyzed from this study consisted of data from 8 students in IN, 13 students in IS, 20 students or 10 dyads in PN, 20 students or 10 dyads in PS, and 12 students or 6 dyads in DYN. Data from a few additional participating students was lost due to technical problems during the data collection.

Results

Table 2 Q7

First we checked the results of Individual versus Pair and System-Support versus no System-Support manipulations to verify whether they were consistent with our first study. The main effect of Pair versus Individual was consistent with the first study. There was significant evidence of process losses when comparing real pairs with nominal pairs $F(1,37)=5.88$, $p<.05$,

Table 2 Summary of results from Study 2

	Individual—No Support	Individual—System Supported	Pair—No Support	Pair—System Supported	Dynamic
Pretest	9.9 (1.8)	10.1 (1.4)	8.8 (1.9)	9.6 (1.4)	9.6 (1.6)
Posttest	10.1 (2.2)	10.4 (1.2)	9.1 (2.0)	9.3 (1.4)	8.9 (1.7)
Unique Ideas from Task1 (individual level analysis)	8.3 (2.7)	9 (3.3)	4.4 (2.5)	5.8 (2.6)	4.8 (2.1)
Unique Ideas from Task 2	4.4 (1.7)	4.2 (1.4)	3.7 (1.9)	4.6 (1.7)	4.6 (1.4)

effect size .65 standard deviations. In the first study, the trend for idea generation to be greater in the System-Supported conditions was not statistically significant except in the group level analysis. However, in this study, the trend was not only in the same direction, but it was significant this time even at the individual level, $F(1,37)=8.9$, $p<.005$, effect size .59 standard deviations. As in the first study, there was no significant interaction between these two factors. Thus, we observe a consistent effect of both manipulations on brainstorming productivity during the experimental manipulation.

Because idea generation performance during the experimental manipulation is consistent with what we observed in the first study, we can compare these results with those obtained in this study for the new, hybrid condition (DYN) to determine whether folk wisdom about brainstorming alone and then as a group does indeed enhance productivity.

At the 5 min mark, we see significant process losses, and no effect of System-Support $F(2, 44)=11.46$, $p<.0001$. A Tukey pairwise post-hoc analysis reveals that nominal pairs are significantly more productive than real pairs (effect size 1.24 standard deviation), whereas the behavior in the hybrid DYN condition is not statistically different from either. It is halfway in between on average. Thus, when students anticipated working with a peer after the end of the first 5 min, they did not experience the productivity advantage of individuals working alone who did not anticipate that. This could potentially be related to the idea of social apprehension, which is one common source of process losses in group brainstorming.

In the final 25 min, we expected students in the hybrid DYN condition to behave like the students in the System-Supported Pairs condition. In fact, the System-Supported Pairs condition performed the best of all conditions in the final 25 min of brainstorming, $F(4,42)=4.1$, $p<.01$. The System Supported Pairs condition performed significantly better than all other conditions according to a students-t post-hoc analysis and better than all but the System-Supported Individuals condition according to a Tukey pairwise post-hoc analysis. According to both post-hoc analyses, contrary to our expectation, the system-supported pairs condition significantly out-performed the hybrid DYN condition, effect size 1.54 standard deviations.

Again, if we examine brainstorming productivity during the entire brainstorming phase summatively, we see that folk wisdom about which brainstorming configuration would be most effective made an incorrect prediction, possibly due to the short duration of the study, and the possibility because the phased structure of the brainstorming session was distracting, and thus disruptive. The overall performance of the hybrid DYN condition is not statistically distinguishable from that in the other conditions, but falls somewhere in the middle. At the five minute point when productivity begins to level off in all conditions, we evidence of an interruption effect in the DYN condition, where the reduction in productivity seems most abrupt. While students seem to recover from this by the 20 min point, their productivity levels off at a lower level on average from the other conditions where students have system support, although as mentioned this is only a statistical trend.

Discussion

While the results from study 2 do not support the new design ideas hypothesized from the results of study 1, they do offer hope that something like the system support offered by the VIBRANT agent has the potential to mitigate process losses that occur because of cognitive interference, especially when the automatic analysis of the discussion that is used to trigger the support is accurate. Note, as mentioned earlier in the paper, the automatic analysis was substantially more accurate in the second study. Results from study 2 confirm that process losses from cognitive interference are a problem even in simple brainstorming tasks such that designs for group brainstorming systems that focus merely on alleviating problems due to production blocking are not sufficient. Furthermore, while we did not observe system support ameliorating process losses when we viewed productivity in the first study, we did observe such an effect in the second study. Again, it is possible that the difference was because in the second study we had the benefit of more accurate conversational analysis technology that was used to trigger the system support. It is possible that the negative effect of system support in the first five minutes in the first study was that it was distracting because of its inaccuracies, and when this was corrected in the second study, it was not a problem. Thus, based on the results of the second study, we argue that brainstorming in groups does have some merit and that in cases where groups are brainstorming,, it is best to offer support for the entire duration of brainstorming. As mentioned above, there was a positive main effect of system support on idea generation productivity, and there was no significant difference in productivity between students who worked in pairs with system support and those who worked individually with system support. Thus, in order to reap all of the benefits of doing idea generation in pairs, the configuration of working in pairs with system support may be the best compromise.

Another important point is that despite the fact that the results of the second study do not confirm the design recommendations hypothesized from the results of the first study, together the studies offer a richer and more comprehensive picture of what is happening in brainstorming groups than either do alone. The second study offers strong evidence that appropriate support can mitigate process losses in brainstorming groups. Reinterpreting the results from the first study in light of the second study highlights the importance of accurate conversation analysis for triggering support.

It is surprising that the hybrid DYN condition turned out to be significantly worse than a condition where participants worked in pairs for the entire duration of the brainstorming, since this contradicts a commonly held belief that it is beneficial to brainstorm alone and then participate in group brainstorming. One possible explanation for the lack of effectiveness of that condition is that it was distracting for students to suddenly be joined by a partner and a feedback agent after 5 min. Nevertheless, if this were the whole explanation, one would expect the effect to wear off after a short time. However, this does not seem to be the case since the idea generation performance of the students in the DYN condition leveled off before it caught up with that of the participants in the condition where they worked in pairs with feedback the entire time.

Another possible explanation is that valuable time was lost when participants were joined by a partner because of having to "catch up" by sharing the ideas that had already generated while working alone. This "catch up" time is typically not considered when the advice is to brainstorm alone before group brainstorming is offered. However, this phenomenon encourages participants to spend time repeating ideas already contributed rather than focusing on generating new ones. It is possible that this act of repeating ideas exacerbates an effect related to cognitive interference (Wang and Rosé 2007), leading students to feel prematurely that they do not have any new ideas to contribute due to over-exposure to already articulated ideas. The results may have looked different if the total time for brainstorming was not so short.

We acknowledge that the support offered by the VIBRANT agent is only feasible in a limited domain, most likely in educational contexts where the same brainstorming task may be used repeatedly because of its educational value. In brainstorming tasks in work contexts, this would not be an option. Thus, as part of our current research we are working towards a more general approach that would allow system support for brainstorming to occur for arbitrary brainstorming tasks.

Future work

While the approach taken in our current system configuration is to offer support based on a topic analysis of conversational contributions, we argue that a more general approach would prompt feedback based on an analysis of the structure of the conversation, in line with current work on automatic collaborative process analysis (Donmez et al. 2005; Wang et al. 2007; Rosé et al. 2008; Mayfield and Rosé 2011), where labels are assigned to conversational contributions according to the role they play within the conversation, and indicate such things as whether participants are building on one another's ideas or talking at cross-purposes. Such an analysis could be used to identify places in the conversation where support is most needed, and to determine where productivity is sufficiently high and it might be better for the system to "back off".

Recall that VIBRANT's feedback is composed of two parts, namely a comment that acknowledges the contributed idea and a tutorial that points the participant towards a new focus for idea generation. Although an automatic process analysis of idea generation that would focus on the dynamics of the conversation rather than the content of specific contributions would not allow us to offer content based feedback, it is not clear that content oriented support offered in the form of VIBRANT's comments acknowledging valid ideas is necessary for stimulating idea generation. We suspect that it is VIBRANT's tutorials, which are based on prior work related to the effect of "category labels" (Dugosh et al. 2000; Nijstad and Stroebe 2006), that are responsible for the positive effect we observed. If support was only offered during low productivity regions of the conversation, and if it was limited to pointing participants to direct their idea generation to particular places within the idea space, it is possible that it would matter less whether the direction of those hints was related to the specific ideas that had already been contributed or not. Further investigation is needed to verify whether this will be a feasible and effective solution.

As part of our long term effort, we are considering other possible directions. Since students in the pairs conditions were observed to sometimes waste time repeating and paraphrasing each other's ideas, one potential future agent design might be one that encourages partners to explore different parts of the idea space widely, but within the same chat space, so that students would have a broad pool of ideas to draw from collaboratively, which may also go some way towards avoiding producing ideas that lead to cognitive interference related process losses in their idea generation (c.f., Nijstad and Stroebe 2006). This design is also consistent with other work related to the jigsaw method (Sharan 1980).

Conclusions

In this paper we present the results of two behavioral studies investigating both the long term and short term effects of brainstorming in pairs versus brainstorming individually in the context of CSCL and inquiry learning. Our finding is that brainstorming tasks can be beneficial for student learning. Furthermore, our data support the view that learning from brainstorming comes from the constructive, inferential process of idea generation building on prior knowledge as well as from the collaborative process of students building on one another's ideas. Beyond that, the results from the first study suggest 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 in terms of connection-based learning 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 in the first study, although with effective support, we did see a positive outcome in the second study that offers hope that with continued development and experimentation with conversational agent based support, we can achieve a successful configuration in which process losses can be mitigated.

Acknowledgements This research was supported by NSF grants HCC-0803482, DRL-0835426, and SBE 0836012.

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), 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., & 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., & 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., Barufaldi, J. P., Lin, M., & Chen, Y. (2007). Assessing tenth-grade students' problem solving ability in the area of Earth sciences. *Computers in Human Behavior, 23*(4), 1971–1981.

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.

Connolly, T. (1993). Behavioral decision theory and group support systems. In L. Jessup & J. Valacich (Eds.), *Group support systems* (pp. 270–280). NY: Macmillan.

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.

Dutoit, A. H. (1996). The role of communication in team-based software engineering projects, PhD Thesis, Computer and Electrical Engineering, Carnegie Mellon University, Pittsburgh PA.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education, 7* (1), 22–30.

Graesser, A. C., Person, N., Harter, D., & TRG. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12*(3).

Gustafson, K., & Bennett, W. (1999). *Issues and difficulties in promoting learner reflection: Results from a three-year study.* www: http://it.coe.uga.edu/~kgustafs/document/promoting.html.

Gweon, G., Rosé, C. P., Carey, R., & Zaiss, Z. S. (2006). Providing support for adaptive scripting in an on-line collaborative learning environment. *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2006).*

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

Kapur, M., & Kinzer, C. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning, 4*, 21–46.

Kraut, R. E. (2003). Applying social psychological theory to the problems of group work. In J. Carroll (Ed.), *Theories in human-computer interaction* (pp. 325–356). New York: Morgan-Kaufmann Publishers.

Kumar, R., Beuth, J., & Rosé, C. P. (2011). Conversational strategies that support idea generation productivity in groups. In: *Proceedings of Computer-Supported Collaborative Learning.*

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (Brainstorming). *European Journal of Social Psychology, 37*, 822–832.

Mayfield, E., & Rosé, C. P. (2011). Recognizing authority in dialogue with an integer linear programming constrained model. In: *Proceedings of the 49*th *Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.*

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.

NIST (2002). The economic impact of inadequate infrastructure for software testing, RTI, Research Triangle Park for the National Institute of Standards & Technology, Report 02–03, May 2002.

Piaget, J. (1985). *Equilibration of cognitive structures*. University of Chicago Press.

Popescu, O., Aleven, V., & Koedinger, K. (2003). A knowledge based approach to understanding students explanations. *Proceedings of the AI in Education Workshop on Tutorial Dialogue Systems: With a view towards the classroom*. IOS Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271.

Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research, 50*(2), 241–271.

Wang, H., & Rosé, C. P. (2007). A Process Analysis of Idea Generation and Failure. *Proceeding of the Annual Meeting of the Cognitive Science Society.*

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., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers in Education, 46*(1), 71–95.

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