CHAPTER 4

Pilot study on game situation for creative thinking

To evaluate the effectiveness of Idea Storm Cube, we divide the evaluation into two parts. The first part aims to evaluate the feasibility of a game-based environment for enhancing Paradigm-Modifying thinking while the second part is to evaluate the mechanism for arousing new ideas. In this chapter, we present the design of the experiment for the first part.

4.1 Pilot design

We adopt a behavior observation approach to investigate the feasibility when students collaboratively think of a problem within a game-type situation. We set our hypothesis as that an interesting game-stimulated environment is useful for students brainstorming. We simplify the treatment by removing the agent support and potential idea valid mechanism.

This pilot was conducted in May 2007, in a computer classroom at Taipei Municipal Jianguo High School. The duration of the experiment was approximately 45 minutes including the Idea Storm Cube game and two tests. There were a total of 16 10th-grade students involved in this experiment. All of them have taken the course of earth science. In this pilot, we made four (an arbitrary number) peers as one group and assumed that this small group size is appropriate for brainstorming. Therefore, there were four groups with four peers in each group. We employed the pilot treatments and tests as below:

Group: R	T_1	S_1	S_2

R = random assignment of students

T = experimental treatments, Idea Storm Cube game

S_1 = Domain-Specific Knowledge Test & Reasoning Skill Test

S_2 = Open-end Creativity Problem-Solving Test

First, when the students logged in the ISC, all participants will be randomly assigned to different groups. By using the random sampling method which is based on the sampling theory in statistic, we had four groups which have almost the same estimate of the population.

Second, we gave the operational instruction in detail and told them that this was a collaborative brainstorming activity, but they did not know who their teammates are. All groups had their own game-competition ISC for brainstorming. Then we try to evaluate the users' behavior when they played within the game-type brainstorming system.

Finally, all groups were instrumented to finish two tests which are the Domain-Specific Knowledge Test (DSKT) and the Reasoning Skill Test (RST) [4] for evaluating their abilities of the domain knowledge and the reasoning skill. Therefore, we can see not only their behavior in ISC, but also whether they have different performances on these two abilities or not, and then examine our hypothesis.

4.2 Data Analysis

In order to observe the participants' behaviors in our game-type brainstorming system, a large amount of action data were recorded. These data have the following features:

- The ideas, submitted times, contributor, and the validity status by the system
- The participant's views when they were browsing the ideas by the rotation actions

Then we use two measurements to evaluate their using behaviors. The first one is the ratio of idea category, with which we can observe the idea diversity that the user generates. The

other one is the percentage of valid ideas *Rv*, which is the ratio of the valid ideas over all ideas. Both of them are the parameters used to adjust the agent support although we did not use agent here.

4.3 Results

After the pilot study, we had obtained some statistics describing the participants' using behaviors. All of the 16 participants submitted 285 ideas, and the mean of their submitted ideas is 17.81 (SD = 7.52) in 15 minutes. Every idea proposed by a participant was labeled into 19 domain concepts in the aforementioned Domain Model (seeing the detail in APPENDIX D) by two independent coders, and the inter-rater reliability was satisfactory (Kappa = .739). The mean of valid idea is 6.37 (SD = 1.58), the mean of invalid idea is 6.56 (SD = 5.83), and the mean of meaningless idea is 6.68 (SD = 3.96). Note that the invalid idea means the idea is not matched with 19 domain concepts and the meaningless idea indicates that the idea is irrelevant responses with the topic of this task. By comparing the valid ideas among these participants as shown in Figure 4-1, we found that some ideas are rather duplicated such as the idea06 and idea07, but the frequency of valid idea is rather different (i.e., some of valid ideas are high, but some do not appear in all participants' responses). Therefore, these data need further analysis for the status of submitting ideas and the diversity of concept.

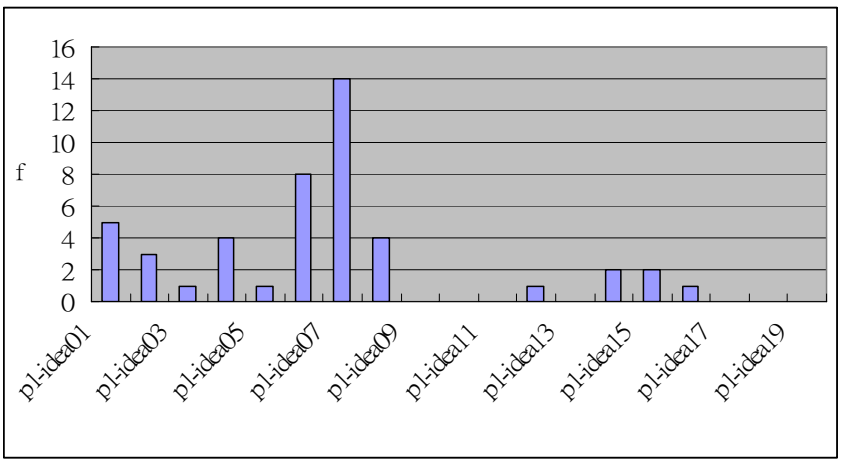

Figure 4-1: The frequency of valid ideas in all participants' responses

First, for their submitting ideas, the R_v (i.e., it is each participant's ratio of valid ideas over all ideas) was computed and the mean of this ratio is 0.52 (i.e., the valid ideas is the majority over all ideas when the ratio > 0.5). It seems that the valid ideas is little more than the invalid ideas. However, if we divide the experiment time, 15 minutes, into three segments of 5 minutes, then we can observe the interesting results as shown in Table 4-1.

Table 4-1: The mean, SD of invalid ideas, valid ideas and the R_v

	The status of submitted ideas		
	Invalid ideas *Mean (SD)*	Valid ideas *Mean (SD)*	R_v *Mean (SD)*
0~ 5 min	0.87 (1.50)	3.37 (1.62)	0.84 (0.21)
5~10 min	2.93 (3.21)	1.81 (1.22)	0.43 (0.30)
10~15 min	3.06 (2.14)	0.93 (0.77)	0.29 (0.28)
ANOVA			
F (2, 45)	4.218	15.435	17.64
p	0.021*	0.000*	0.000*
Effect size			
η^2	0.158	0.407	0.439
f	0.433‡	0.828‡	0.885‡

* $p < 0.05$ Effect size: ‡large

Rv is the ratio of the valid ideas over all ideas

The means of R_v decrease as the time goes on as shown in Figure 4-2. It makes sense because people usually propose all ideas as soon as possible in a brainstorming activity and may exhaust his ideas at a later time. Observably, this problem is not what we aim to solve in ISC of this pilot study.

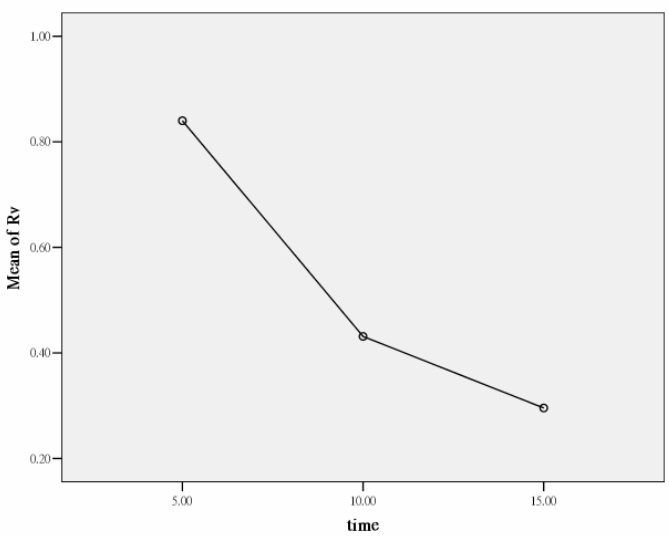

Figure 4-2: The ratio of valid idea decreases at the time goes on

This is the so called *idea blocking* phenomenon (i.e., it is also described as a production blocking in other researches [15][23]) that we want to overcome in a brainstorming activity. This phenomenon also occurs in individual thinking or collaborative discussing in group. In our design, the agent support and the incremental validating mechanism of potential ideas are used to relief this bottleneck as we will see in the result of the next experiment.

As for the ratio of idea category, the diversity of the participant's ideas is explored. The 19 ideas in the domain model have been divided into 5 categories. We also classify all participants' ideas into these categories. Table 4-2 shows the average of the number of ideas by each participant, the expected number of ideas (i.e. original idea count in each category) and the idea coverage of each category in this pilot study. Although we cannot compare the coverage of these categories directly, the numbers show that an uneven distribution exists among these categories.

Table 4-2: The count average, expected count, and coverage of each category

	Geology	**Ecology**	**Nature**	**Development**	**Policy**
count average	1.31	1.69	0.50	2.63	0.00
expected count	3.00	2.00	1.00	10.00	3.00
coverage	0.43	0.84	0.50	0.26	0.00

This result can be explained as that the participants' knowledge for this science topic is homogeneous, so they generate the same ideas in certain popular category. Another probable reason is that the idea exchange mechanism by rotating the cube makes the participants have homogeneous ideas. In terms of knowledge learning, it is not a bad news for students to share their understanding of the domain. However, how to increase the lower coverage of some categories as shown in Figure 2-1 is an important issue for us because the system aims to inspire divergent thinking. In other words, the system may make the students learn more Paradigm-Modifying way of thinking if we can encourage them to generate ideas in the lower categories.

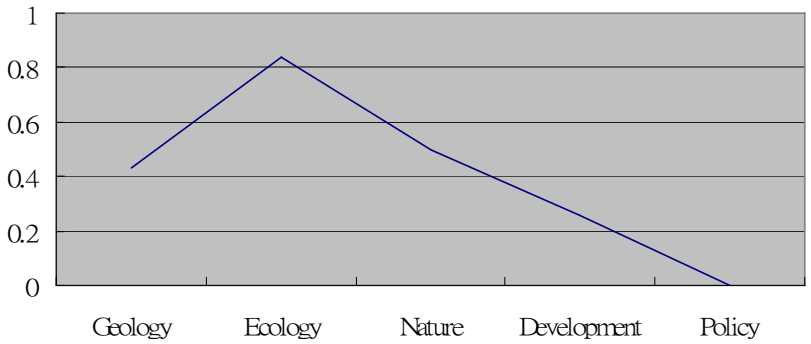

Figure 4-3: The coverage of each category

Without the control group in this pilot study, we did not know whether this result was good or not. Nevertheless it can be compared with an information-type GSS or agent-based ISC in the next experiment.