CHAPTER 5
Main Study

After the pilot study, we have conducted the second part of the study on evaluating the incremental mechanism that arouses new ideas and on comparing the effects of system support among game- and agent-based ISC.

5.1 Participants in the experiment

This experiment was conducted in June 2007, in a computer classroom at Taipei Municipal Nanhu High School. The duration of the experiment was approximately 45 minutes including the pre-test, different group treatments, and post-test. There were a total of 72 11th-grade students involved in this experiment. All of them have taken an introductory course on earth science.

5.2 Experimental design

We adopted a different brainstorming type of comparison-group approach to investigate the comparative effects that students worked collaboratively on an idea generation problem either within an Information-based, Game-based, or Game&Agent-based condition, and their function comparison as Table 5-1 shown.

Table 5-1: The comparison of three groups

	Info-based ISC	Game-based ISC	Game&Agent-based ISC
Information-sharing window	√		
Game competition rules		√	√
Agent support			√

Hence, we set our hypothesis as:

H1: An interesting game-based environment is more useful than Information-based support system for brainstorming in terms of concentration enhancement on a task.

H2: During the brainstorming process, a peer-like agent can be made more helpful than human peers by giving supports from domain knowledge and thus avoid the problem of idea blocking.

Therefore, we employed the experimental treatments and tests as below:

Group 1 (Infomation-based ISC):	R	S_1	T_1	S_2
Group 2 (Game-based ISC):	R	S_1	T_2	S_2
Group 3 (Game&Agent-based ISC):	R	S_1	T_3	S_2

R = random assignment of students

S1 = Domain-Specific Knowledge Test & Reasoning Skill Test (pre-test)

T1, T2, T3 = experimental treatments (T1 = ISC_{info}, T2 = ISC_{game}, T3=$ISC_{game\text{-}agent}$)

S2 = Open-end Creativity Problem-Solving Test (transfer test)

We will view the different versions of the system being used as the independent manipulation and observe whether there is a significant difference for the outcomes at the individual level and the group level among the three groups.

First, when the students logged in the ISC, every three human peers were grouped into a single ISC triad, and all participants had been randomly assigned to the group with the conditions of ISC_{info} (4 groups), ISC_{game} (10 groups), or $ISC_{game\text{-}agent}$ (10 groups). The sample sizes are different because we only have limited subjects in this experiment. We also had a pre-test (15 minutes) of the Domain-Specific Knowledge Test (DSKT) and the Reasoning Skill Test (RST) [4] to measure student's knowledge of the science topic in order to ensure that we have the similar estimate of the population. Therefore, there are three groups with almost the same knowledge for the task to run the experiments on the three different brainstorming-support systems.

Second, we had different experimental treatments (15 minutes) for the three groups. All of them were ask to do a brainstorming task. Here, we use the Debris Flow Hazard (DFH) task which has been designed by science educators [3] in the area of Earth sciences. The ISC_{game} and the $IS\text{-}C_{game\text{-}agent}$ groups both implemented game-competition rules for brainstorming. An additional intelligent agent that monitors the users' inputs and responds to them with stimulating ideas was available in the $ISC_{game\text{-}agent}$ group. The ISC_{info} group had an information-sharing ISC without game rules for exchanging cards. All of these three systems immediately give credits to the users and show scores on their screens if they propose good ideas that are validated by the system. But there is no rotation suggestion in the ISC_{info} group. By comparing ISC_{info} to ISC_{game}, we hope to see whether the game-type brainstorming may inspire more paradigm-modifying ideas or not, and by $ISC_{game\text{-}agent}$, we hope to see whether the presence of a peer-like agent in addition to the game-rules would even makes students have more paradigm-modifying thinking.

Finally, three groups were instrumented to finish the third test (15 minutes) which is the open-end Creativity Problem-Solving (CPS) Test [20]. It aims to observe the quality and quantity of the paradigm-modifying ideas that students produce. At the end, we want to observe whether the students in different groups have different performance on the divergent thinking ability in order to examine our hypothesis.

5.3 Brainstorming with different supportive mechanisms

Although the three groups have the brainstorming activity on the same topic, the mechanism they used in each group is different.

The ISC_{info} group has an information-sharing window on the top-right of the screen to show each idea submitted from the group teammates as Figure 5-1 shows. The student can scroll the text window to see all past ideas in his group but he does not get credit if he writes the same idea that has appeared in this window.

Figure 5-1: An information-sharing window in ISC$_{info}$ group

The ISC$_{game}$ group and ISC$_{game\text{-}agent}$ group have the same representation without the information-sharing window as shown in Figure 5-2. The student can see his teammates' ideas only if he or someone rotates the cube, i.e. he only has a limited view on the ideas. Similarly, no credit is given if he writes the same idea as his teammates.

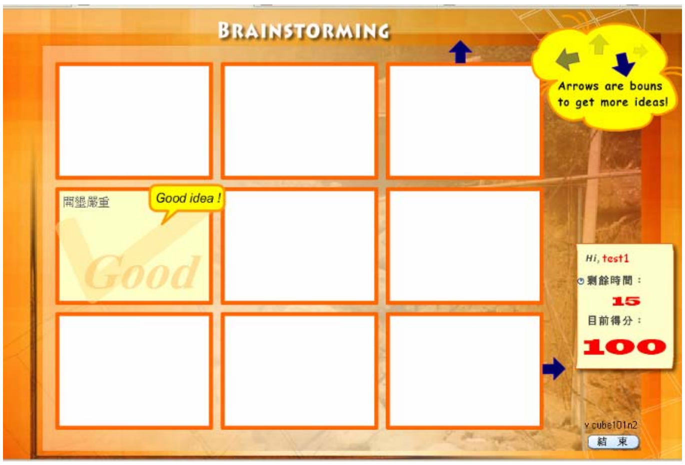

Figure 5-2: The same representation without the information-shring windows in ISC$_{game}$ group and ISC$_{game\text{-}agent}$ group

Moreover, a special intervention in ISC$_{game\text{-}agent}$ group is that a peer-like agent will present and play with users in the brainstorming activities. From the result of the prior pilot study, we know that students might encounter idea blocking after the first five minutes. Therefore, it is reasonable for the peer-like agent to give support when they are likely to encounter this problem. In this experiment, we postpone the start time of the agent to 5 minute time after the game starts in order to let students generate their own ideas first. The presentation of the ideas from the agent relies on cube rotations, i.e. those artificial peer-like ideas only show up when someone or the agent rotates the cube (referring to the agent rotating mechanism in 3.6.2 section). Note that the participants did not know the existence of peer-like agents.

5.4 Measurement

In order to examine our hypothesis, we need to compare the effect of the game stimulation between the ISC_{info} group and the ISC_{game} group, and the support effect of the peer-like agent between the ISC_{game} group and the $ISC_{game\text{-}agent}$ group. We set some indices to compare those groups for observing students' Paradigm-Modifying ability in terms of different type of brainstorming mechanisms.

For the game stimulation, we concern about the students' aspiration and whether they have the right thinking direction when they play this brainstorming game. To measure this, we record the number of each student's submitted ideas. That is, how many new potential ideas have been submitted, and how many ideas were valid according to our Domain Model.

In addition, we consider that peer-like agents should be helpful because students may encounter idea blocking. Hence, we need measurement for the group-level outcome. Thus, we use group as the evaluation unit. We would like to know the coverage of different idea category and the number of unique ideas that each group proposed.

We also have a transfer test that asks students the questions such as "*what facilities or solutions may prevent a debris flow hazard from happening?*" for observing the Paradigm-Modifying ability performance on similar science problem after they finished these brainstorming activities. Therefore, this experiment used the following variables as predictors to predict the students' Paradigm-Modifying outcome:

In the individual-level outcome:

- The number of submitted ideas: n_{submit}
- The number of new potential ideas (i.e. were judged as potential ideas by the system and may be considered as good ideas by human coder in post audit) : n_{new}
- The number of valid ideas: n_{valid}
- The coverage of idea categories (i.e. the number of valid ideas in each idea category): $c_{category}$
- The transfer test score of the open-end Creativity Problem-Solving test: $s_{transfer}$

In group-level outcome:

- The number of unique valid ideas in a group: gn_{unique}
- The number of unique valid ideas of each group in each category: $gc_{category}$

The comparison with the measurements the Torrance Test of Creative Thinking (TTCT) [42] are showed as below:

Table 5-2: The comparison of the measurements between ISC and TTCT

Idea Stroming Cube	**Torrance Tests of Creative Thinking [42]**
Individual-level	
The number of new potential ideas	Originality (statistical rarity of the responses)
The number of valid ideas	Fluency (total number of meaningful ideas)
The coverage of idea categories	Flexibility (the number of different categories)
Group-level	
The number of unique valid ideas	Fluency (total number of meaningful ideas)
The number of unique valid ideas of each category	Flexibility (the number of different categories)

* The measurement ‘elaboration’ did not be used here

Note that we just try to present the possible meaning of these measurements that we suggested by comparing with another measurement in common use. In addition, the number of submitted ideas is used as the measurement of attention and the transfer test score is taken as the measurement for transfer effect in a new situation.

5.5 Data analysis

An ANOVA was computed on the pre-test data and found that all of them are not significant (*p* >.05) among three groups. All of 54 participants (i.e., because some participants were due to the loss of system failure) submitted 1660 ideas and the mean of their submitted ideas was 30.74 (SD = 13.16), which is almost twice as the pilot study.

As in the pilot study, each participant's ideas were labeled into 19 domain concepts in the aforementioned Domain Model. In order to validate the coding results with limited research resources, we employed a SVM classification model trained by coding results labeled previously in another study [14][46] to perform automatic coding. The performance of the SVM classification model has known to achieve reliable level (Kappa = .7). By this automatic coding software, the inter-rater reliability between one independent coder and the automatic coding software was very close to satisfactory (Kappa = .668).

In order to examine and compare the effect of these three groups, an analysis of variance (ANOVA) was conducted on those measurements as mentioned in the last section. Thus, the independent variable of the analysis was the type of supportive mechanism (i.e., Information-based, Game-based, Game&Agent-based), and the dependent variables were those measurements mentioned in the previous section.

Several tests we made to ensure that the assumptions were met in ANOVA for this experiment. The Kolmogorov-Smirnov tests show that the number of submitted ideas (*p* =1.000), the number of new potential ideas (*p* =.371), the number of valid ideas (*p* =.349), and the number of unique ideas in group (*p* =.534) were all normal distributions (*p* >.05). The Levene's tests of the homogeneity of variance also show that the error variance for each dependent variable is equal across groups. Therefore, the results of assumption tests permit the use of the analysis of variance for this experiment. The analysis was undertaken in SPSS 13.0 (Statistical Package for Social Sciences version 13).

5.6 Results

5.6.1 Individual-level outcome

Table 5-3 shows the descriptive statistics of participants' ideas generation statistics assessed by different types of brainstorming groups. As mentioned before, the time period that each participant played was of the same length (15 minutes) but from the descriptive statistics we can roughly observe that there were large difference means among groups in n_{submit}, and less difference means in n_{valid}.

Table 5-3: Descriptive statistics of participants' the number of submitted ideas, valid ideas, and new potential ideas among the ISC_{info} group, ISC_{game} group, and $ISC_{game\text{-}agent}$ group

Treatment	Submitted ideas (n_{submit}) *Mean (SD)*	Valid ideas (n_{valid}) *Mean (SD)*	Potential ideas (n_{new}) *Mean (SD)*
ISC_{info} (n=12)	20.17 (8.85)	6.42 (2.15)	2.83 (1.85)
ISC_{game} (n=24)	33.58 (11.80)	7.71 (1.37)	4.00 (2.54)
$ISC_{game\text{-}agent}$ (n=18)	34.00 (14.15)	8.94 (2.26)	4.72 (3.41)
ANOVA			
F (2, 50)	5.900	6.583	1.710
p	0.005*	0.003*	0.191
Effect size			
η^2	0.188	0.205	0.063
f	0.481‡	0.507‡	0.259†

* *p* < 0.05   Effect size: ‡large, †medium

We further conducted ANOVA tests on those dependent variables as shown in Table 5-3. The results revealed that for n_{submit}, F (2, 50) = 5.90, p =.005, and for n_{valid}, F (2, 50) = 6.583, p =.003, showing significant differences (p <.05). On the other hand, this study reports not only statistical significance test, but also practical significance (effect magnitudes). The effect size index *eta squared* (η^2) or f was used because it is more appropriate for the analysis of variance or covariance [7]. According to Cohen's rough characterization (1988, pp. 284-288), f = 0.1 is deemed as a small effect size, f = 0.25 a medium effect size, and f = 0.4 as the large effect size (for interpreting η^2, 0.01 = "small," 0.059 = "medium," and 0.138 = "large"). The effect size (η^2 = 0.188 or f = 0.481 in n_{submit}, η^2 = 0.205 or f = 0.507 in n_{valid}) according to the criteria of effect size proposed by Cohen [7] is close to a large effect size ($f > 0.4$) and it gives us an estimate of the noteworthiness of the result as McLean [27] said. As for n_{new}, there were no significant differences among groups, but medium effect size ($\eta^2 = 0.063$ or $f = 0.259$).

However, we still cannot tell the differences between each pair of conditions among the three conditions. Thus, we need to run a Post Hoc multiple pairwise comparison of Scheffe test (i.e., Scheffe test for all contrasts in multiple comparisons is the most frequently testing method when the number of participants are different across conditions) to identify which condition outperformed.

Table 5-4: The results of Scheffe test for the number of submitted ideas and valid ideas.

Scheffe	Submitted ideas (n_{submit})			Valid ideas (n_{valid})		
	ISC_{info} *Mean (SD)*	ISC_{game} *Mean (SD)*	*p*	ISC_{info} *Mean (SD)*	$ISC_{game-agent}$ *Mean (SD)*	*p*
	20.17 (8.85)	33.58 (11.80)	0.011	6.42 (2.15)	8.94 (2.26)	0.003
	ISC_{info} *Mean (SD)*	$ISC_{game-agent}$ *Mean (SD)*	*p*			
	20.17 (8.85)	34.00 (14.15)	0.013			

Table 5-4 shows the results of Scheffe test. For the number of submitted ideas, there was a significant difference between two pairs of groups which were ISC_{info}-ISC_{game} ($p = .011 < .05$) and ISC_{info}-$ISC_{game-agent}$ ($p = .013 < .05$), but no significant difference between ISC_{game} and $ISC_{game-agent}$. This result indicates that the game competition within brainstorming seems better for participants to generate and write ideas than information-sharing type of brainstorming.

For n_{valid}, there was also a significant difference between ISC_{info} and $ISC_{game-agent}$ ($p = .003$), but no significant difference between other pair of groups. This result shows that the brainstorming context of game competition with support from the peer-like agent made participants generate more ideas in correct thinking directions than the information-sharing type of brainstorming. As for the number of new potential ideas, there was no any significant difference among each pair of groups.

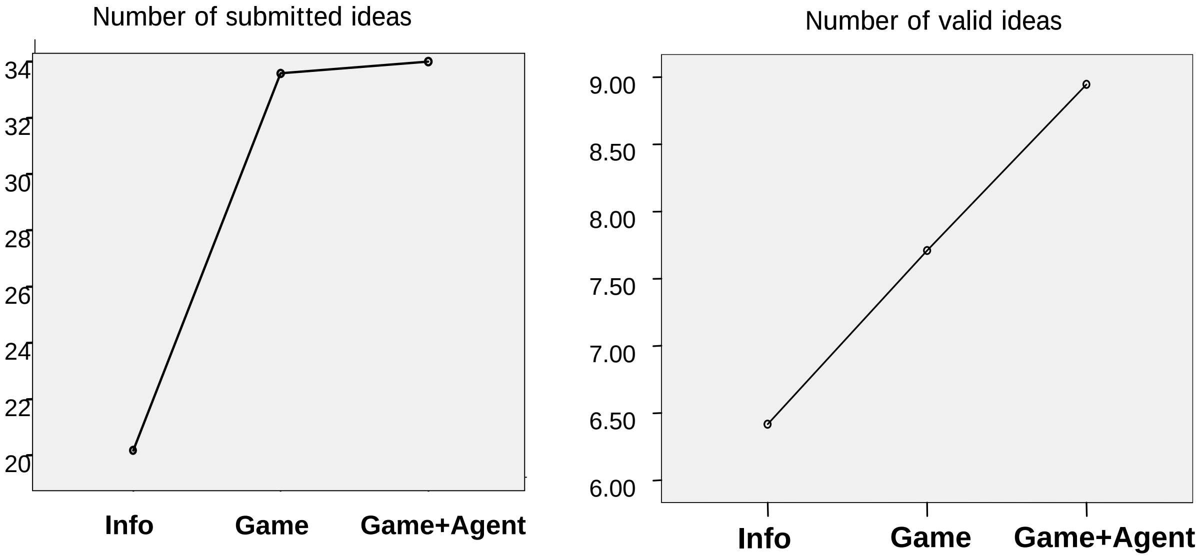

Figure 5-3: The plotting figures of the number of submitted ideas and valid ideas

Preliminary findings from Figure 5-3 are that the game rules significantly improved attempts of generating ideas, and the peer-like agent further improved productivity and efficiency of idea generation. The reason of efficiency from the agent is that the number of submitted idea is almost the same between ISC_{game} and $ISC_{game\text{-}agent}$, but $ISC_{game\text{-}agent}$ is better than ISC_{game} in the number of valid ideas.

Moreover, in order to understand the diversity of participants' ideas, we looked into $c_{category}$ which is the number of valid ideas falling into five categories under this domain (i.e. geology, ecological environment, natural factor, artificial development and policy). Table 5-5 presents the descriptive statistics of ideas falling into each specific category. We further run ANOVA and Scheffe test to evaluate the difference among various experimental conditions.

Table 5-5: Descriptive statistics of each participants' coverage of each category, number of participants: ISC_{info} group (n=12), ISC_{game} group (n=24), and $ISC_{game\text{-}agent}$ group (n=18).

Treatment	Geoglogy *Mean (SD)*	Ecology *Mean (SD)*	Nature *Mean (SD)*	Human Development *Mean (SD)*	Policy *Mean (SD)*
ISC_{info} (n=12)	1.75 (0.62)	1.67 (0.65)	0.50 (0.52)	2.42 (1.24)	0.08 (0.28)

ISC$_{game}$ (n=24)	2.04 (0.75)	1.79 (0.41)	0.63 (0.49)	3.54 (1.10)	0.17 (0.38)
ISC$_{game-agent}$ (n=18)	2.11 (0.83)	1.67 (0.68)	0.67 (0.48)	3.50 (1.24)	0.94 (0.63)
ANOVA					
F (2, 50)	0.892	0.320	0.421	4.094	17.865
p	0.416	0.728	0.659	0.022*	0.000*
Effect size					
η^2	0.034	0.012	0.016	0.138	0.412
f	0.188†	0.110†	0.128†	0.400‡	0.837‡

* *p* < 0.05 Effect size: ‡large, †small

From the results in Table 5-5, we can see that the categories of geology (p = .041), ecology (p = .000), and policy (p = .000) were significantly different, but no significant difference was found in the ecology and nature topics.

Table 5-6: The results of Scheffe test for the categories of human development and policy.

Scheffe	Human Development			Policy		
	ISC$_{info}$ *Mean (SD)*	ISC$_{game}$ *Mean (SD)*	*p*	ISC$_{info}$ *Mean (SD)*	ISC$_{game-agent}$ *Mean (SD)*	*p*
	2.42 (1.24)	3.54 (1.10)	0.034	0.08 (0.28)	0.94 (0.63)	0.000
				ISC$_{game}$ *Mean (SD)*	ISC$_{game-agent}$ *Mean (SD)*	*p*
				0.17 (0.38)	0.94 (0.63)	0.000

We further take a multiple comparison of Scheffe test to identify which contrast possesses statistical significance. The results of Scheffe test, as Table 5-6 shows, revealed that the ISC$_{info}$ condition and the ISC$_{game-agent}$ condition differed in all the three categroies. In addition, the ISC$_{info}$- ISC$_{game}$ was also significant in artificial development category and the ISC$_{info}$- ISC$_{game}$ was also significant in the policy category.

This result shows that the three brainstorming-support systems had significant different effects for individuals to generate ideas on different categories. Especially in the comparison of ISC$_{info}$ and ISC$_{game-agent}$, we can easily find the supporting effect on generating divergent ideas because the performance of the ISC$_{game-agent}$ condition is better than the ISC$_{info}$ condition in many categories. This indicated that brainstorming in the content of game competition with support from peer-like agent made participants generate ideas on more different aspects than information-sharing type brainstorming system. However, with regard to the pairs of ISC_{info}-ISC_{game} and ISC_{game}-$ISC_{game\text{-}agent}$, we only can find few supporting effect on generating divergent ideas because the performance is better in only one category.

For individual outcomes, we summarize the significant results as below:

Measurement	Performance	*p*
Submitted idea	ISC_{game}, $ISC_{game\text{-}agent}$ > ISC_{Info}	.01
New potential ideas	No sig.	.
Valid ideas	$ISC_{game\text{-}agent}$ > ISC_{Info}	.00
The coverage of idea categories	$ISC_{game\text{-}agent}$ > ISC_{game} > ISC_{Info}	.03

5.6.2 Group-level outcome

For the group-level outcome, we want to evaluate gn_{unique}, which is the unique ideas, and $gc_{category}$, which is the ideas coverage of different categories, for the groups. First, we see the descriptive statistics of gn_{unique} among three conditions as shown in Table 5-7 and we found that it was significant difference (p =.000). Therefore, we also needs more elaborated ANOVA and Scheffe tests.

Table 5-7: Descriptive statistics of the number of unique ideas from each group: ISC_{info} group, ISC_{game} group, and $ISC_{game\text{-}agent}$ group

Treatment	Unique Ideas of Group (gn_{unique}) *Mean (SD)*
ISC_{info} (n=4)	10.25 (1.89)
ISC_{game} (n=8)	12.12 (0.64)
$ISC_{game\text{-}agent}$ (n=6)	14.16 (0.40)
ANOVA	
F (2, 15)	19.618
p	0.000*
Effect size	
η^2	0.723
f	1.616‡

* $p < 0.05$ Effect size: ‡large

Table 5-7 shows the results of ANOVA, and it reveals that gn_{unique} in group, F (2, 15) = 19.618, p =.000, has significant differences and the effect size (η^2 = 0.723 or f = 1.616) is a pretty large effect size. Then we run the Post Hoc multiple comparisons of Scheffe test (n is different) to identify which pair-group has significant difference.

Table 5-8: The results of Scheffe test for the number of unique ideas of group.

Scheffe	Unique ideas of Group		
	ISC_{info} *Mean (SD)*	ISC_{game} *Mean (SD)*	*p*
	10.25 (1.89)	12.12 (0.64)	0.024
	ISC_{game} *Mean (SD)*	$ISC_{game\text{-}agent}$ *Mean (SD)*	*p*
	12.12 (0.64)	14.16 (0.40)	0.006
	ISC_{info} *Mean (SD)*	$ISC_{game\text{-}agent}$ *Mean (SD)*	*p*
	10.25 (1.89)	14.16 (0.40)	0.000

Table 5-8 shows the results of Scheffe test, and it indicates that the number of group un-duplicated ideas has significant differences among three groups (p = .024, p =.006, p =.000). This result explains that the ISC_{game} group is more helpful for generating different ideas than the ISC_{info} group and the $ISC_{game\text{-}agent}$ group is even better than the ISC_{game} group as shown in Figure 5-4.

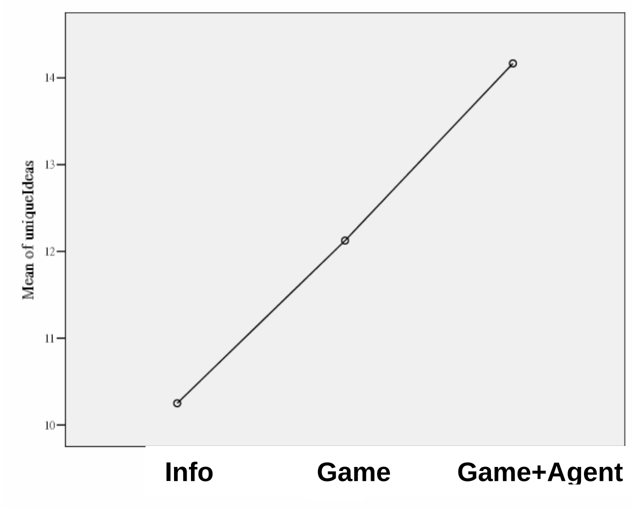

Figure 5-4: The plotting figure of the mean of unique valid idea in each group

Second, we want to see the $gn_{category}$ group idea coverage of the categories. Table 5-9 shows the descriptive statistics of unique ideas count in a group in each category and their ANOVA analysis and Scheffe test.

Table 5-9: Descriptive statistics of the group idea coverage in each category. Number of groups: ISC_{info} group (n=4), ISC_{game} group (n=8), and $ISC_{game\text{-}agent}$ group (n=6)

Treatment	Geoglogy *Mean (SD)*	Ecology *Mean (SD)*	Nature *Mean (SD)*	Human Development *Mean (SD)*	Policy *Mean (SD)*
ISC_{info} (n=4)	3.00 (0.00)	2.00 (0.00)	1.00 (0.00)	4.00 (1.41)	0.25 (0.50)
ISC_{game} (n=8)	2.75 (0.46)	2.00 (0.00)	1.00 (0.00)	5.88 (0.99)	0.50 (0.53)
$ISC_{game\text{-}agent}$ (n=6)	2.83 (0.40)	2.00 (0.00)	1.00 (0.00)	6.50 (0.54)	1.83 (0.75)
ANOVA					
F (2, 50)	0.536	.	.	8.152	11.007
p	0.596	.	.	0.004*	0.001*
Effect size					
η^2	0.067	.	.	0.521	0.595
f	0.268†	.	.	1.043‡	1.212‡

* $p < 0.05$ Effect size: ‡large, †medium

Table 5-9 shows the results of ANOVA, and it reveals that the number of unique ideas in artificial development, F (2, 15) = 8.152, p =.004, and policy, F (2, 15) = 11.007, p =.001, were significantly different. Then we ran the Post Hoc multiple comparisons of Scheffe test (n is different) to identify which pair-group has significant difference.

Table 5-10: The results of Scheffe test for the count of unique ideas in Human Development and Policy

Scheffe	Human Development			Policy		
	ISC_{info} *Mean (SD)*	ISC_{game} *Mean (SD)*	*p*	ISC_{info} *Mean (SD)*	$ISC_{game\text{-}agent}$ *Mean (SD)*	*p*
	4.00 (1.41)	5.88 (0.99)	0.023	0.25 (0.50)	1.83 (0.75)	0.004
	ISC_{info} *Mean (SD)*	$ISC_{game\text{-}agent}$ *Mean (SD)*	*p*	ISC_{game} *Mean (SD)*	$ISC_{game\text{-}agent}$ *Mean (SD)*	*p*
	4.00 (1.41)	6.50 (0.54)	0.005	0.50 (0.53)	1.83 (0.75)	0.004

Table 5-10 shows the results of Scheffe test, and it indicates that there were significant differences in both categories (p = .005 in artificial development, p =.004 in policy) between the ISC_{info} condition and the $ISC_{game\text{-}agent}$ condition. In addition, there were also significant difference in the artificial development category (p = .023) between the ISC_{info} condition and the ISC_{game} condition and in the policy category (p = .004) between the ISC_{game} condition and the $ISC_{game\text{-}agent}$ condition.

This result explains that the ISC_{game} group is more helpful in generating different ideas than the ISC_{info} group and the $ISC_{game\text{-}agent}$ group is even better than the ISC_{game} group. These group outcomes almost match the result of the supporting effect in the outcome at the individual-level.

For group outcomes, we summarize the significant results as below:

Measurement	**Performance**	***p***
Valid ideas	$ISC_{game\text{-}agent}$ > ISC_{game} > ISC_{Info}	.02
The coverage of idea categories	$ISC_{game\text{-}agent}$ > ISC_{game} > ISC_{Info}	.02

5.6.3 Transfer test

After the different treatment, we adopted a transfer test in order to know whether the students can transfer what they learn to improve their thinking ability in solving new problems. The transfer test is the paper-based test as shown in APPENDIX C and the question of the task is similar with the main task of the experiment, but without any computer support such as information-sharing, game, or agent support.

The result is a kind of learning outcome and different from the process outcomes in 5.6.1 and 5.6.2 sections. Without any computer support and gaming environment, it becomes very different from the main task in learning process. Table 5-11 shows that the results of the transfer test and it revealed that the transfer test were significant differences (p = 0.033) and effect size is close to large (f = 0.387).

Table 5-11: Descriptive statistics, ANOVA and effect size of the transfer test

Treatment	Unique Ideas of Group *Mean (SD)*
ISC_{info} (n=12)	35.75 (8.65)
ISC_{game} (n=22)	27.04 (11.50)
$ISC_{game\text{-}agent}$ (n=18)	33.11 (7.82)
ANOVA	
F (2, 49)	3.661
p	0.033*
Effect size	
η^2	0.130
f	0.387‡

* p < 0.05 Effect size: ‡ large (close to 0.4)

Then we run the Post Hoc multiple comparisons of Scheffe test (n is different) to identify which pair-group has significant difference as shown in Table 5-12. This result reveals that the ISC_{info} group has higher performance than other groups, especially ISC_{game} group (p = 0.054, close to significant difference), in the transfer test. It is worthy to note that this result is contrary to some results of the main task in this experiment.

Table 5-12: The results of Scheffe test for the count of unique ideas in Human Development and Policy

Scheffe	Transfer test		
	ISC_{info} *Mean (SD)*	ISC_{game} *Mean (SD)*	*p*
	35.75 (8.65)	27.04 (11.50)	0.054

5.7 Discussion

As we know, creativity is very difficult to measure and evaluate although there are several creativity testing scales [42]. In this experiment, we observed many using behaviors in different types of brainstorming-support systems and evaluated what kind of system support could help students have more diversity thinking. It deserved further discussions on whether the divergent thinking equals creativity or not. Here, we consider that this ability is positive to creativity and only focus on the measurement of the divergent thinking ability.

We cited Bostrom's measurement of Paradigm-Relatedness [2] to evaluate participant's divergent thinking behavior. Therefore, we calculated their valid ideas in different category because we want to observe the diversity of thinking about different view of points for the task. That is our rationale why the artificial system can enhance students' divergent thinking ability by using the systematic method of gaming stimulation or adaptive support.

Before discussing the outcomes of the brainstorming system, an important thing that should be confirmed first is the user's intention for brainstorming. Different attitude sometimes will lead different results, especially for mental activities such as brainstorming. Hence, it is the goal of the brainstorming-support system to let the user enjoy the process and concentrate on thinking for the task and generating novel ideas.

In our experimental hypothesis H1, we assumed that an interesting game-based environment is more useful than Information-based brainstorming-support system. The major reason is that the game competition may enhance students' concentration on their brainstorming task. Relating it back to the result of this experiment, if the number of submitted ideas could indicate the users' intention to their brainstorming task, we can find that the hypothesis is valid because the number of submitted ideas and the number of valid ideas were significant differences between the different types of brainstorming system with/without game rules. In addition, from the participants' open-end opinions and feedbacks as shown in APPENDIX E, we know that most participants were interested in this activity.

Another point to be elaborated here is about the limited view of information. As we just mentioned, the bounded rationality [36] of thinking make people have cognitive load and only focus on sub-topic of the total task. We turn this limitation into a chance to conform humanity by designing user interface for the limited information and a game rule for 'exchanging' information.

We may infer that the brainstorming-support system with game competition rules and appropriate user interface design makes participants have higher intention to 'play' this brainstorming task. That also matches Csikszentmihalyi's flow theory [6]. Although we know that the game competition is not always suitable for every student, especially for the students with lower learning achievement, it is suggested that the game-type is good for most students.

Regarding the divergent thinking ability, from the experimental results, we want to see the effect of the peer-like agent support with game competition rules. In our experimental hypothesis H2, we consider that the presence of a peer-like agent is more helpful than the presence of human peers because collaborative discussion may encounter idea blocking that we mentioned in the pilot study. The peer-like agent could give supports from domain knowledge. In our experimental results, there are some evidences to support this idea in individual or group as shown in Table 5-13.

Table 5-13: The significant different results of the pair ISC_{game} - $ISC_{game\text{-}agent}$

	Mean (SD)	*p*
Individual outcomes		
The coverage of idea categories	0.17 (0.38) - 0.94(0.63) in Policy	.000
Group outcomes		
The number of unique ideas	12.12 (0.64)- 14.16 (0.40)	.006

The coverage of idea categories in Policy	0.25 (0.50) - 1.83 (0.75)	.004

Therefore, we can see that the game-based brainstorming-support system with a peer-like agent has better performance comparing with the game-based brainstorming-support system. We may infer that there will be good performance on divergent thinking by enhancing user's intention to exchange ideas with his teammates and giving some effective support when they encounter idea blocking. This result matches the Csikszentmihalyi's model of creativity [5] which is combining individual's idea, social opinion and expert's knowledge.

For other results form this experiment, the correlation between pre-test and transfer test very low (Pearson's r is 0.215). It may result from the different testing types between two tests (i.e., one is knowledge test and the other is idea generation test) or the treatments that we manipulated in different ISC groups. It is a pity that we do not have a formal post-test to measure the variability of participants' knowledge, and thus we cannot make any inference here, especially for the interesting result that ISC_{info} is better than ISC_{game} in the transfer test score (p = 0.05). Therefore, we think whether the design of limited view would affect the working memory in cognitive load because the ISC_{info} groups have the information-sharing window to review their all ideas and maybe there was some learning effects from repeating browsing information. Or the thinking ability is hard to change only by playing this activity once. We cannot know the effect in this experiment and it deserves more analysis of these results. In addition, the number of potential new ideas was set as the originality of TTCT [42] in this experiment, but we did not see any significant differences here. We use the rule of potential idea discovering in ISC_{game} and $ISC_{game\text{-}agent}$ groups but there were not effect by the game rule. One possible reason is that most ideas of the Debris Flow Hazard task have been collected into the Domain Model fUP_d, or another possible reason is that the heuristic method of idea validation is not good enough to recognize those responses among participants. This is also worthy to take more explorations for evaluating the effect of this special game rule.

In this research, we do not have the comparison of idea generation between individual and group in this game-based support system. However, many studies [29] show that the productivity loss is even greater in groups than in individuals. It is suggested that the individual version of this system should be considered in further studies.

Summarizing the result obtained in this experiment, we can see the effect from the game competition rules and the peer-like agent support by comparing three different types of conditions. It is suggested that the game-based brainstorming system with appropriate intelligent support is applicable, especially if we are going to use this system for students to learn the knowledge of different subjects because this brainstorming-support system can make them think more from diverse view points for learning the knowledge.