Original article

Triple scheme of learning support design for scientific discovery learning based on computer simulation: experimental research

Jianwei Zhang*, **Qi Chen**†, **Yanquing Sun**† **& David J. Reid**‡
*Educational Technology Center, Tsinghua University, Beijing, China
†School of Psychology, Beijing Normal University, Beijing, China
‡School of Education, University of Manchester, Manchester, UK

Abstract Learning support studies involving simulation-based scientific discovery learning have tended to adopt an *ad hoc* strategies-oriented approach in which the support strategies are typically pre-specified according to learners' difficulties in particular activities. This article proposes a more integrated approach, a triple scheme for learning support design on the basis of the systematic analysis of the internal conditions of scientific discovery learning. The triple learning support scheme involves: (a) interpretative support that helps learners with knowledge access and the generation of meaningful and integrative understandings; (b) experimental support that scaffolds learners in systematic and valid experimental activities; and (c) reflective support that increases learners' self-awareness of the discovery processes and prompts their reflective abstraction and integration. Two experiments were conducted with eighth graders (13-year-olds) to examine the effects of these learning supports embedded into a simulation program on floating and sinking. The overall results support the main hypotheses that learning supports in a simulation environment should be directed towards the three perspectives to invite meaningful, systematic, and reflective discovery learning.

Keywords learning environment, learning support, scientific discovery learning, simulations

Introduction

In the past decade, research on discovery learning has evolved from concept discovery learning (CDL) towards more sophisticated and authentic scientific discovery learning (SDL) characterized by the necessity of designing scientific experiments (van Joolingen & de Jong 1997). Since computer simulations have the capacity to provide learners with an exploratory learning environment, they are regarded as powerful tools for SDL. However, many of the researchers, who have compared the effects of simulation-based learning to more traditional modes of learning, have found little persuasive evidence of their effectiveness (see de Jong & van Joolingen 1998; Lee 1999). The following question arises: why does simulation-based learning that involves learners in active inquiry not improve learning outcomes more consistently? One explanation lies in the wide range of difficulties learners may have in dealing with discovery learning tasks. De Jong and van Joolingen (1998) classified the difficulties that learners may encounter into four categories: (a) difficulties in generating and adapting hypotheses, (b) poorly designed experiments, (c) difficulties in data interpretation, and (d) problems regarding the regulation of discovery learning. Despite its potential in stimulating constructive learning activities, the

Accepted: 1 September 2003
Correspondence: Jianwei Zhang, Educational Technology Center, Tsinghua University, Beijing 100084, China.
Email: zhangjw@tsinghua.edu.cn

simulation-based learning environment cannot guarantee effective learning without sufficient support (scaffolding) for discovery learning activities.

In order to promote effective discovery learning, a number of studies have been conducted to help learners accomplish particular tasks during the learning processes. For example, some researchers developed supportive methods to help generate hypotheses in simulation-based discovery learning (Shute & Glaser 1990; Njoo & de Jong 1993; Quinn & Alessi 1994). Others have looked at the issues connected with experimental design (Leutner 1993), planning (Tabak *et al*., 1996), and access to an appropriate knowledge base (Lewis *et al*. 1993). So far, most studies of simulation-based discovery learning have adopted an *ad hoc* support strategies-oriented approach. This approach proposes specific support strategies according to learners' difficulties in particular aspects in terms of the effects of the proposed learning supports. No study was found to have made a systematic analysis of the internal conditions that determine the effectiveness of SDL or depicted a scheme for learning support designed in the light of the internal conditions.

SDL is a typical form of constructive learning based on problem solving activities involving the design and implementation of scientific experiments. Scientific discovery is usually interpreted as the processes of mindful coordination between hypothesized theories and evidence collected by experiments (Kuhn *et al*. 1992; Klahr & Dunbar 1988). SDL is a knowledge construction approach that is accomplished based on scientific discovery activities. Three main interlocked spheres exist in the processes of effective SDL (see Zhang 2000): (1) problem representation and hypothesis generation, which heavily relies on the activating and mapping of prior knowledge and the meaning-making activities, (2) testing hypotheses with valid experiments, and (3) reflective abstraction and integration of the discovery experiences. Taking all these perspectives into account, we hypothesized that three interrelated conditions may determine the effectiveness of SDL. These are:

1 *The meaningfulness of discovery processes*: Learners need to activate their prior knowledge and map that knowledge onto the problem being addressed to help in representing the problem and generating appropriate hypotheses and understandings.

2 *The systematic and logical discovery activities*: Effective discovery learning involves proper scientific reasoning, systematic manipulations of the variables, and valid designs and implementations of experiments.

3 *The reflective generalization* over the discovery processes, which means the self-monitoring of the discovery processes and the reflective abstraction and integration of the discovered rules and principles.

The four categories of difficulties that learners may encounter during SDL, which have been summarized by de Jong and van Joolingen (1998), can all be attributed to the limitations in these three conditions. According to the three hypothesized conditions, three types of learning support can be designed and geared towards the three spheres:

(a) *interpretative support (IS)* that helps learners with knowledge access and activation, the generation of appropriate hypotheses, and the construction of coherent understandings;

(b) *experimental support (ES)* that scaffolds learners in the systematic and logical design of scientific experiments, the prediction and observation of outcomes, and the drawing of reasonable conclusions; and

(c) *reflective support (RS)* that increases learners' self-awareness of the learning processes and prompts their reflective abstraction and integration of their discoveries.

So far, most studies designed to support SDL focus on the impacts of certain specific support strategies, most of which are directed towards systematic and logical experiment and discovery activities (e.g. Rivers & Vockell 1987; Njoo & de Jong 1993; Tabak *et al*. 1996). More studies need to be carried out to examine the effects of the three types of learning support and propose a comprehensive scheme for learning support design.

Within the triple scheme of learning support design for simulation-based SDL, an experimental study was conducted to investigate the effects of two types of learning support embedded in a simulation environment: ES and IS (Reid *et al*. 2003). In the results, IS manifested prominent main effects on the post-tests of intuitive understanding, flexible application, and knowledge integration. However, there was no significant effect for ES on the post-tests or the quality of learners' experiments conducted during discovery processes. ES in this study included a number of elements such as the explanations about experimental design (especially 'varying one thing at a time'), the prompts for identifying the objective of each experiment, the prompts for predicting and observing outcomes, and drawing conclusions. However, these treatments were still not supportive enough to improve learners' experimental activities or learning outcomes. This result disagrees with Rivers and Vockell's (1987) finding that providing learners with general experimentation hints before their exploration could promote their experimentation abilities, as well as Swaak *et al.*'s (1998) conclusion that experimental support in the form of assignments had a clear effect on SDL.

The aim of the present research was to further examine the effects of ES and IS, and in addition to explore the effect of the RS. Study I investigated the effects of ES and IS on simulation-based discovery learning among learners with different reasoning abilities. Study II focused on the effect of RS on simulation-based discovery learning. In our previous study (Reid *et al.* 2003), as one of the treatments of ES, the program merely gave learners a general explanation about experimental design (e.g. 'You'd better vary one factor at a time, otherwise you cannot make clear which factor is having an effect'). In Study I, this treatment was improved by exemplifying the principle of experimental design with a specific example. In addition, some treatments of ES in the form of questions were changed from selective to compulsory ones. Also, in order to investigate the influence of learners' reasoning ability, Study I included learners' reasoning ability levels as a between-subjects factor in the experiment design.

Study I: Effects of ES and IS on simulation-based scientific discovery learning

Methodology

Simulation-based learning environment

The topic chosen for the simulation was floating and sinking, where the students were required to explore the upthrusts (buoyant forces) upon objects either submerged or floating in water. This topic is one of the core concepts in secondary science learning and encompasses the representative variable structure of scientific discovery tasks. The English version of the simulation is shown in Fig. 1. The essential rule governing the size of an upthrust is Archimedes' law of buoyancy, stating that the size of the upthrust acting on an object in liquid equals to the weight of the displaced liquid. This is the essential physics rule for objects that either float or sink. However, it is difficult for young learners to discover this rule together with the underpinning displacement method in a short time. Especially, it is even harder for them to think about the amount of liquid displaced by a floating object in comparison to a sinking object. Therefore, rather than asking the students to discover this essential rule (the ultimate full model) of upthrusts, this study designed a progressive task (de Jong *et al.*, 1999) for them to explore the relationship between the features of an object and the size of its upthrust in the circumstances of floating and sinking. The students were required to complete two tasks by manipulating this simulation software: (1) Task F: to discover which one or more of the three given factors (shape, mass, and volume) were related to the size of the upthrust acting upon a *floating* object; (2) Task S: to discover which one or more of the three given factors were related to the size of the upthrust on a *submerged* object. It can be argued that the two tasks are comparable in difficulty because they have similar problem structure. Learners often hold misconceptions about this phenomenon, e.g. assuming that the size of the upthrust depends on the shape or volume of the floating object. Actually, for objects submerged in water, it is the volume of object that influences the size of the upthrust, because the volume of the object determines the volume of the displaced water, which affects the size of the upthrust. For a floating object, the size of upthrust always equals to the size of its weight. These two forces (weight and upthrust) balance each other and result in the equilibrium of the object in water. So the mass of an object is the only feature that is related to the size of upthrust under this circumstance. Although by nature, Archimedes' law can explain these variable relationships under both floating and sinking circumstances, it should be noted that Archimedes' law was beyond the discovery tasks in this study.

The simulation adopted paired-instance design that required learners to construct a pair of experiments at a time, so that they could contrast the outcomes of two instances conveniently. For example, in order to examine the effect of the volume of an object, learners dragged two objects of the same shape (e.g. a ball) to the top of the left and right containers, set the values on the left and right to keep the masses the same, and varied the volumes (see Fig. 1(a)). Then they clicked the 'RUN' button to see whether the upthrusts were different or not. For all the students, a data sheet was provided on screen to record and display the value of

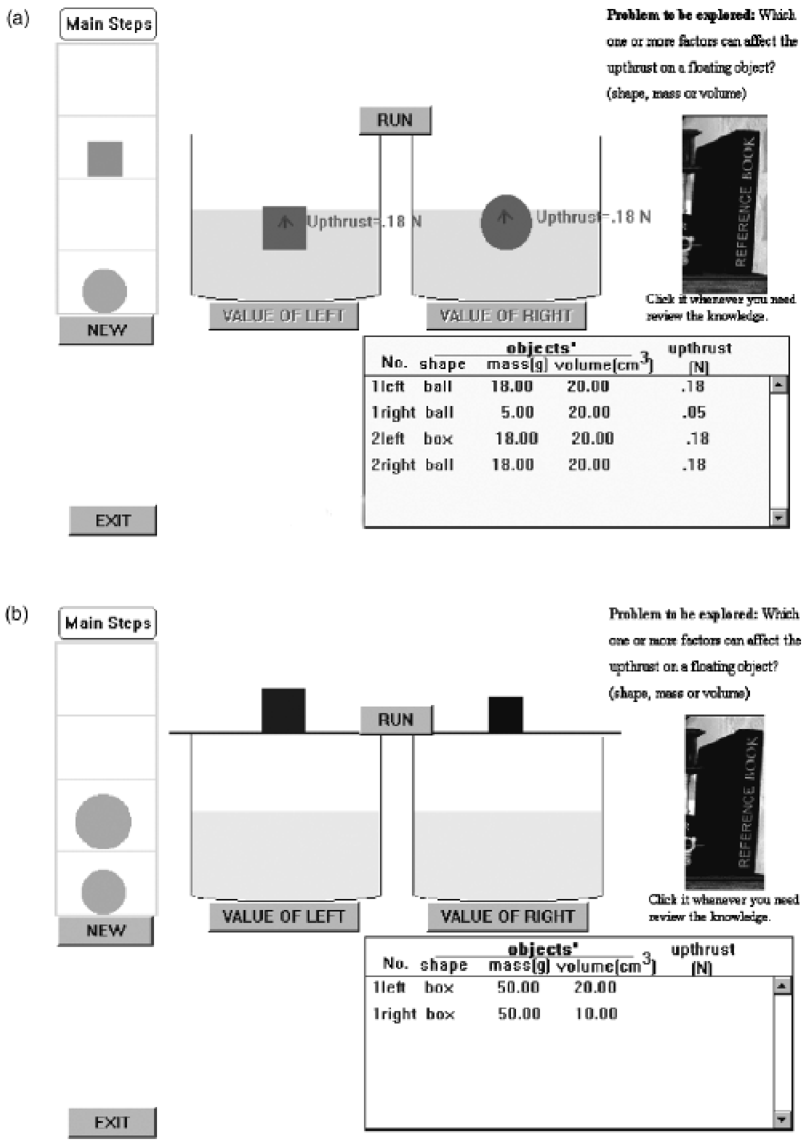


Fig. 1 Interface of the simulation (a) during Task F (after clicking the 'RUN' button) and (b) during Task S (before clicking the 'RUN' button).

the input and output variables in each pair of experiments. In addition, a permanent button 'Main Steps' was prepared to remind learners of the main steps in an experiment, which involved selecting and dragging objects, setting their values, running the experiments, observing outcomes, and clicking the 'NEW' button to start another trial.

The learning environment contained two kinds of learning supports: ES and IS. The ES, intended to help learners conduct valid experiments, included four specific treatments. These were:

(a) *Explanations about scientific experimental design.* In the introductory phase of the simulation, the program gave learners general explanations about scientific experimental design (particularly about 'varying one thing at a time') and exemplified the ideas with a specific case of experiment design (exploring which factors are related to how soon sugar dissolves in water).
(b) *Identify the objectives of each trial.* Before designing each pair of experiments, learners were required to identify their objectives by ticking the variable(s) (shape, mass, and volume) they wanted to examine.
(c) *Prediction and comparison.* Learners were required to predict which of the two specified objects would have a larger upthrust before running the

experiments, and to check their predictions against the outcomes after the experiments.

(d) *Conclusion.* After each pair of experiments, they were asked to conclude their new discoveries against an experiment structure table showing the comparisons of the input and output variables between the two objects in a pair of experiments (see Appendix I).

Since the effect of IS was examined as a within-subjects factor in this study, IS was only present in Task F and not in Task S. IS, which was intended to support learners from the perspective of meaning-making, consisted of three measures:

(a) *Activating prior knowledge.* A multiple-choice question was offered in the introductory section of the simulation program to activate learners' prior knowledge about balanced forces, which asked: 'For an object *floating* in water, which force or forces are acting on it?'

(b) *General analysis of the problem.* Three multiple-choice questions (without feedback) were provided prior to the discovery processes in order to prompt the students to make a general analysis of the factor(s) that are relevant to the size of the upthrust on a floating object. Hmelo and Day (1999) ever found that embedding questions in the simulation could function as an anchor to help students connect simulation closely to scientific understandings.

(c) *Access to a knowledge base.* A permanent button (*reference book*), which contained descriptions of the concepts of weight, balanced forces, motion, as well as the basic meaning of 'upthrust', helped learners access relevant knowledge during the discovery learning process.

The simulation program was written in such a way that it registered learners' manipulations during the learning processes and wrote a log-file for each learner.

Research design

In order to investigate the effects of ES and IS among learners with different reasoning abilities, a 2 (ES/no ES) $\times$ 2 (IS/no IS) $\times$ 3 (high/middle/low ability) mixed design was adopted. ES and reasoning ability were included as between-subjects factors, and IS was involved as a within-subjects factor. As noted previously, every learner was asked to accomplish two discovery tasks: Task F (upthrust on floating objects) that encompassed IS and Task S (upthrust on submerged objects) without IS. The two tasks were presented by computer randomly in order to balance the possible impact of the sequence of the two tasks. Log-files were used to analyse how learners had processed their discovery learning.

Participants

The participants were 80 eighth graders from a junior high school in urban Beijing. The students were 13-year-olds on average. In their science classes, they had been taught about the relevant concepts of mass, force, balanced forces, motion, as well as the basic meaning of upthrust, but not the exact rules being addressed in this study. All the participants had attended an introductory course of information technology and had the basic computer skills that were necessary to manipulate the simulation software in this study. In order to explore the impact of learners' reasoning ability on discovery learning, the Raven's Standard Progressive Matrices test was chosen to evaluate the participants' reasoning ability. The range of mean ± 0.5 standard deviation was defined as the middle-ability group, and above and below constituted the high-ability and low-ability group, respectively. The participants of each level were randomly assigned to two conditions: ES or no ES. Table 1 gives the numbers of participants in each group.

Post-tests

In order to gauge the effects of the learning supports on various aspects and levels of discovery learning outcomes, three types of post-tests were written for Task F and Task S, respectively. (1) *The Principle Knowledge Test* assessed how learners had discovered and memorized the basic rules for Task F and Task S using multiple-choice items. (2) *The Intuitive Understanding Test* measured learners' intuitive understandings (instead of fact memorization) about the phenomena, which is regarded as an important goal in SDL (de Jong *et al.* 1999; Swaak & de Jong 2001), especially when conceptual changes are involved.

Table 1. Distribution of the participants in each group

	Low ability	Middle ability	High ability
No ES	13	17	11
ES	11	16	12

ES = experimental support.

Using pictures, these items showed pairs of objects with different combinations of shapes, masses, and volumes and asked learners to predict how their upthrusts would compare in size (see Appendix II for an instance). (3) *The Flexible Application Test* intended to determine how well learners could transfer the knowledge to new problem situations (e.g. analysing the upthrust acting upon a floating boat). These questions were more flexible, requiring the transfer of the discovered principles and the integration of learners' prior knowledge.

Procedure

All the students were given the Raven's Standard Progressive Matrices test 2 weeks before the experiment. The formal experiment took place in a computer laboratory equipped with 50 networked Pentium computers. The participants were required to finish the following sessions individually:

- *Warm-up*. The participants worked with a tutorial version of the simulation program. Three researchers were present to answer questions regarding the operations of the software. This stage lasted approximately 10 min.
- *Discovery and post-tests*. Each participant was required to accomplish two discovery tasks by manipulating the simulation software: Task F and Task S. Each discovery task was allotted 30 minutes, followed by the corresponding post-tests in written format. In order to balance the possible impact of the sequence of Task F and Task S, the computer program randomized the sequence of the two tasks. Participants were clearly reminded that their task was to discover the correct rules on the basis of sufficient evidence through simulated experiments.

Results

Effects of the ES, IS and reasoning ability on the post-tests

Principle knowledge. In MANOVA using ES and reasoning ability level as the between-subjects factors, and IS (Task F/S) as the within-subjects factor, a marginally significant effect was found for IS ($F(1, 74) = 3.57$, $P < 0.10$),[1] showing that learners had gained higher scores in Task F that encompassed IS compared with Task S that had no IS (Table 2). A significant main effect was observed for reasoning ability on the principle knowledge test ($F(2, 74) = 11.00$, $P < 0.001$), indicating that learners with higher reasoning ability could accomplish the discovery task more successfully. There was no significant main effect for ES. However, a notable interaction was observed between ES and reasoning ability ($F(2, 74) = 3.95$, $P < .05$) (see Fig. 2). Simple effect analysis revealed that ES had a marginally significant positive effect among low-ability participants ($F(1, 76) = 2.96$, $P < 0.10$), a significant negative effect among middle-ability participants ($F(1, 76) = 4.59$, $P < 0.05$), and no notable influence among participants with high reasoning ability ($P > 0.10$).

Intuitive understanding. There was a significant main effect with reasoning ability ($F(2, 74) = 7.59$, $P < 0.01$), showing that participants with higher reasoning ability scored higher on this test. ES manifested a significant negative effect ($F(1, 74) = 5.13$, $P < 0.05$). Participants who had received ES gained lower scores on the intuitive understanding test.

Flexible application. A prominent significant effect was observed for IS ($F(1, 74) = 91.70$, $P < 0.001$), indicating that learners had gained much higher scores in Task F with IS than in Task S without IS.

Evaluation of learners' experiments

Quality of learners' experiments. Using the data collected in the log-files, an analysis was made to scrutinize how the participants had designed their experiments with the simulation program. 'Change one thing at a time' is an important principle in scientific experiment. Unfortunately, learners are often found to vary many variables in one experiment (Glaser *et al*. 1992; de Jong & van Joolingen 1998). Focusing on the principle of variable control in scientific experiment, we used the following three indices to evaluate learners' experiment designs.

Index I – Percent of well-controlled experiments. As was mentioned in the methodology section, this simulation used paired-instance design that required learners to conduct a pair of experiments at a time, so that they could contrast the outcomes of two instances conveniently. Index I indicates the percentage of the well-controlled experiments in which one and only one factor (shape, mass or volume) was kept different

[1] Since the sample in each cell was relatively small, marginally significant effects at the level of 0.10 will be reported to prevent possible ignorance of meaningful patterns.

Table 2. Means and SDs of the three types of post-tests on Task F and Task S

Groups		Task F (with IS)						Task S (without IS)					
		Principle knowledge		Intuitive understanding		Flexible application		Principle knowledge		Intuitive understanding		Flexible application	
		M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
High ability	ES	6.63	0.70	4.50	0.80	5.04	1.42	6.33	0.98	4.17	1.11	3.42	1.56
	No ES	6.73	0.65	4.82	0.60	4.73	1.62	6.73	0.90	4.82	0.40	3.36	1.29
Middle ability	ES	5.06	2.46	3.56	1.50	4.63	1.64	4.44	2.94	3.88	1.36	3.33	1.06
	No ES	6.53	1.00	4.71	0.77	5.48	1.56	5.94	2.30	4.29	1.45	3.53	1.23
Low ability	ES	4.82	2.48	3.45	1.21	4.82	0.98	4.73	2.87	3.55	1.57	3.00	1.33
	No ES	4.31	2.59	3.69	1.60	4.79	1.82	2.85	2.91	3.54	1.27	3.17	1.40

IS = interpretative support, ES = experimental support.

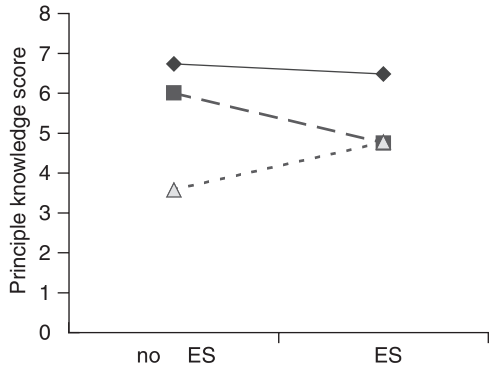


Fig. 2 Interaction of ES and reasoning ability on the principle knowledge test.

between the left and right side instances (see Fig. 1(a) as an example).

Index II – Average number of variables varied in each pair of experiments. This is a looser criterion indicating how many variables among shape, mass, and volume were varied in each pair of experiments on average.

Index III – Focused examination of the three variables. We identified a pair of experiments as having undergone a 'focused examination' of certain variables (shape, mass, or volume) if that variable was the only variable varied in the pair of experiments. For each variable, a full score of 2 was given when it had been examined by at least two pairs of experiments at different levels of the controlled variables (an example is shown in Appendix III). Score 1 indicated that the variable had been examined by only one pair of experiments or by more than one pair of experiments, but at the constant levels of controlled variables. Sequentially, score 0 meant that no experiment had been focused on this variable at all. An average score across the three variables was used in the final analysis. Among the three indices, Index III is the strictest one in that it is the only index that indicates the distribution of well-controlled experiments across the three variables. Learners' scores on the three indices are shown in Table 3.

MANOVAs were implemented for the three indices using ES and reasoning ability level as the between-subjects factors and IS as the within-subjects factor. On Index I, ES had a significant main effect ($F(1, 65) = 3.99$, $P = 0.05$), showing that participants with ES had conducted a larger proportion of well-controlled experiments than those without ES. There was a marginally significant main effect for reasoning ability

Table 3. Means and SDs of the indices evaluating learners' experiments

Groups		Task F (with IS)						Task S (without IS)					
		Index I		Index II		Index III		Index I		Index II		Index III	
		M (%)	SD	*M*	SD	*M*	SD	*M* (%)	SD	*M*	SD	*M*	SD
High ability	ES	60.41	0.18	1.50	0.30	1.27	0.42	48.26	0.15	1.61	0.23	1.03	0.25
	No ES	49.67	0.22	1.69	0.41	1.24	0.52	53.70	0.22	1.59	0.35	1.30	0.41
Middle ability	ES	47.53	0.18	1.65	0.27	0.95	0.41	52.34	0.18	1.58	0.23	0.90	0.50
	No ES	39.22	0.21	1.72	0.30	0.92	0.45	43.51	0.19	1.76	0.31	0.98	0.43
Low ability	ES	41.11	0.29	1.73	0.45	0.85	0.60	58.42	0.25	1.52	0.42	0.82	0.43
	No ES	31.94	0.20	1.97	0.32	0.77	0.44	36.24	0.21	2.00	0.47	0.81	0.41

Index I: percent of well-controlled experiments in which only one variable was varied.
Index II: average number of variables changed in each pair of experiments (maximum = 3).
Index III: focused examination of the three variables (maximum = 3).
IS = interpretative support, ES = experimental support.

Table 4. Correlation between the quality of experiments and the post-tests

	Task F			Task S		
	Index I	Index II	Index III	Index I	Index II	Index III
Principle knowledge	0.18	– 0.30***	0.24**	0.30***	– 0.40****	0.42****
Intuitive understanding	0.07	– 0.09	0.07	0.22*	– 0.22*	0.35***
Flexible application	– 0.12	– 0.05	– 0.14	0.04	– 0.05	0.11

*$P<0.10$, **$P<0.05$, ***$P<0.01$, ****$P<0.001$.

($F(2, 65) = 2.77$, $P < 0.10$), indicating that higher ability learners exhibited better performance on this index. On Index II, ES displayed a significant main effect ($F(1, 65) = 6.42$, $P < 0.05$), showing that participants with ES had varied fewer variables in each pair of experiments on average. There was also a marginally significant main effect for reasoning ability level ($F(2, 65) = 2.77$, $P < 0.10$). On Index III, the only significant main effect was observed for reasoning ability ($F(2, 65) = 6.18$, $P < 0.01$). Multiple comparison (LSD) demonstrated that the high-ability group had performed better investigations surrounding these factors than the middle- and low-ability groups.

Correlation between the quality of experiments and the post-tests. In order to explore the relationship between the quality of experiments and the outcomes of discovery learning, Pearson correlation analyses were performed between each of the indices and the post-tests (Table 4). Significant correlation was observed between the three indices and the tests of principle knowledge and intuitive understanding.

Study II: Effect of RS on simulation-based discovery learning

Methodology

Simulation-based learning environment

On the basis of the previous investigations of IS and ES, Study II continued to use the simulation software about upthrust to examine the effect of RS on simulation-based discovery learning. In this experiment, the participants were only required to complete Task F, that is, to discover which one or more of three given factors (shape, mass, and volume) were related to the size of the upthrust on an object floating in water.

In this study, all the participants received ES and IS as the common basic learning support. On top of these, the participants in the experimental group were provided with RS, which included:

1 *Tracing and displaying learners' discovery processes in detail*. For learners in the experimental group, the data sheet on the screen would not only record and show the values of the input and output

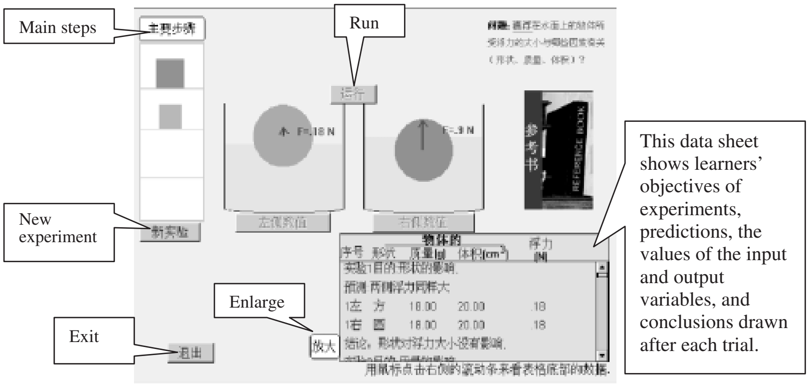


Fig. 3 Interface of the simulation with RS.

variables, but also the objectives of experiments, the predictions before each trial, and the conclusions drawn afterwards (see Fig. 3). All the information was provided to raise learners' self-awareness and monitoring of the discovery processes.

2 *Reflection notes*. After each trial in experiments, learners with RS were required to fill a form that guided them to reflect over the experiment, especially the contrasts between their predictions and the actual outcomes.

3 *Final summarization and generalization*. After a learner clicked the 'Exit' button and finished all the experiments, (s)he would be asked to fill a final summarization and generalization form, which aided learners to reflect over the primary thoughts in designing scientific experiments, the processes and methods in examining the impact of shape, mass and volume, and the explanations of the final conclusions.

Research design

One-way between-subjects design was used to compare the discovery learning outcomes of the RS and No RS groups. Since no pre-test had been implemented for this study, we collected learners' scores in the recent mid-term examination of physics, which took place 1 month before the experiment. This variable was included as the covariate to control the influence of learners' prior academic performance. The log-files generated by the program were used to analyze how learners had accomplished the discovery task.

Participants

The participants were 30 students of eighth grade from the same school as in Study I. The students were 13-year-olds on average. They were randomly assigned to two conditions: RS ($N = 14$) or No RS ($N = 16$).

Post-tests

The principle knowledge test, intuitive understanding test, and flexible application test were also used in this study to evaluate the effectiveness of discovery learning. Moreover, considering the possible effects of reflection on other perspectives of learning outcomes, such as knowledge integration and variable control skills development which have been addressed in the reviewed researches (e.g. Lin & Lehman, 1999), two more categories of post-tests were written for this study:

1 *The knowledge integration test*: Two types of items were developed to assess the associations between the discovered rules and learners' prior knowledge, especially the key concept of balanced forces. The first type took the form of 'association-rating' questions that are one of the response formats in the concept map test (Ruiz-Primo & Shavelson 1996). Learners were required to indicate whether each of the provided concepts was related to their understanding about upthrust or not, and to give a brief justification if they responded with 'YES'. Two concepts including 'weight' and 'balanced forces' were set as target items, and three other concepts were included as filler items. Both learners' ticking and justifications were used to decide the associations in their knowledge. The second style of items involved four 'instance-clustering' tasks (for a similar method, see Chi *et al*. 1981). A picture showing a beach ball floating in a swimming pool was set as the prototype instance. The other four instances were displayed to ask the learners to identify if each one was similar to the prototype instance and explain why they thought so (see Appendix IV). The knowledge association was inferred according to whether they clustered the instances on the basis of the deep structures (force

Table 5. Means and SD's of the post-tests

	Principle knowledge		Intuitive understanding		Flexible application		Knowledge integration		Variable control skills	
	M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
RS	6.00	1.52	4.14	1.10	6.26	1.30	3.21	1.87	11.79	0.43
No RS	6.50	1.32	4.38	1.20	6.19	1.11	1.75	1.48	10.56	2.61
Overall	6.27	1.41	4.27	1.14	6.22	1.18	2.43	1.87	11.13	2.00

RS = reflective support.

pattern) or merely in terms of their surface features (e.g. moving or stationary, being in water, or on a desk). Since the test of knowledge integration is relatively new in format, reliability analysis was conducted particularly for this test, which revealed an acceptable internal consistency coefficient for this category, $\alpha = 0.69$.

2 *The variable control skill test*: It evaluated learners' variable control strategies in scientific experiment by providing the cases of experiment designs (either well-controlled or with flaws) for learners to make critical evaluations and justifications. Similar methods have been frequently used in the researches of variable control skills (see Ross 1988). Reliability analyses revealed an acceptable internal consistency coefficient, $\alpha = 0.89$.

Results

Effect of RS on the post-tests

ANOVAs were implemented for the post-tests using RS as the between-subjects factor and the recent physics scores as the covariate (Table 5). RS demonstrated significant main effects on the scores of knowledge integration ($F(1, 27) = 4.59$, $P < 0.05$) and variable control skills ($F(1, 27) = 4.03$, $P = 0.05$). Learners with RS outperformed the No RS group on both of the tests.

Evaluation of learners' experiments

Learners' experiments conducted during the discovery processes were evaluated using the three indices noted in Study I (Table 6).

Looking at the descriptive statistic data, the RS group learners surpassed the No RS group on all the three indices. However, ANOVAs on the three indices revealed no significant effect for RS, $F(1, 27) = 1.39$, $P = 0.13$ on Index I, $F(1, 27) = 2.73$, $P = 0.11$ on Index II, and $F(1, 27) = 0.30$, $P = 0.59$ on Index III. Learners' physics scores as the covariate displayed significant or marginally significant effects on all the three indices, $F(1, 27) = 4.44$, $P < 0.05$ on Index I, $F(1, 27) = 3.33$, $P < 0.10$ on Index II, and $F(1, 27) = 3.13$, $P < 0.10$ on Index III.

Table 6. Evaluation of learners' experiments

	Index I		Index II		Index III	
	M (%)	SD	*M*	SD	*M*	SD
RS	66.86	0.22	1.36	0.25	1.29	0.52
No RS	57.87	0.22	1.49	0.31	1.25	0.45
Overall	62.07	0.21	1.43	0.29	1.27	0.47

RS = reflective support.

Discussion

Until now most studies on supporting simulation-based scientific discovery learning have adopted an *ad hoc* strategies-oriented solution. These studies have tended to examine the effect of proposed strategies according to learners' difficulties in particular aspects. The present research was dedicated to make more comprehensive and structural studies to make clear the internal conditions of SDL and depict a scheme for learning support design. On the basis of the theoretical analysis of SDL processes, we hypothesized that three interrelated main conditions may determine the effectiveness of SDL: the meaningfulness of discovery processes, the systematicity and logicality of discovery activities, and the reflective generalization over the discovery processes. Accordingly, three types of learning supports were designed: *IS*, *ES*, and *RS*, the effects of which had been examined by the two experiments reported in this article.

Overall, the function of IS was demonstrated as highly significant in Study I as well as in our previous experiment (Reid *et al*. 2003). IS was designed to support discovery learning by activating the relevant knowledge in learners' memories, facilitating the problem representation and hypothesis generation on the basis of prior background knowledge, eliciting more explanation activities towards the experiments, and promoting the access of a knowledge base. This has been verified to be helpful for learners to construct more elaborate, coherent, and transferable understandings about the explored domain, demonstrating notable effects on the principle knowledge and flexible application tests. All the results support one of our major assumptions in the study: the meaningfulness of discovery activities plays a critical role in effective SDL, and hence should be one of the key targets for instructional support. As Ausubel *et al*. (1978, pp. 519–564) argued, the effect of discovery learning depends on the meaningfulness of the discovery experience. SDL does not end with the discovery of one or two pieces of rule, but is intended to incorporate the findings into learners' profound, elaborate, and coherent knowledge structures, and help learners develop their own 'ideas' on scientific phenomena. To a large extent, these learning outcomes will rely on learners' explanatory and interpretative meaning-making activities (e.g. problem representation, hypothesis generation, and adaptation). Apart from the access and activation of relevant knowledge, which has already been addressed in some studies (Leutner 1993; de Jong *et al*. 1999), instructional supports in this sphere need to do more to foster learners' self-explanation and interpretative activities.

ES included such treatments as explaining and exemplifying the principles and strategies in designing scientific experiment, prompting learners to decide the factor(s) to be examined in each trial, to predict and check the outcomes, and to draw conclusions from the experiments. The effect of ES was verified to a greater extent in this study than in our previous research (Reid *et al*. 2003). As the process analysis tells us, the ES groups outperformed the No ES groups on two of the three indices evaluating their experiments, which were closely correlated to learners' achievements on the post-tests. Students with ES outperformed their counterparts in designing well-controlled experiments.

The effect of ES on the post-tests was reflected to be quite complex in its interaction with reasoning ability level on the principle knowledge test. Whether ES was present or not, students with high reasoning ability could accomplish the discovery learning task quite well. They could develop proper strategies (e.g. controlling extraneous variables in experiments) for their experiments relying on their own reasoning ability. ES tended to be most helpful for students with low reasoning ability. In a study, Veenman and Elshout (1995) found that the structural guidance for discovery processes had effective influence on those learners who had relatively lower ability and were inefficient in discovery learning strategies. The present study converges with their finding on this point. Contrary to our expectation, a negative effect was observed for ES among the middle band of learners. A possible reason might be that ES had caused extra cognitive load or distracted learners from their regular thinking processes. Some of the ES treatments took the form of questions for learners to answer; for instance, ticking the factors to be examined, ticking one's predictions, checking the predictions, and drawing conclusions. All the tasks might bring about extra cognitive load for learners and interrupt their regular thinking processes. For learners with low reasoning ability, the discovery task was just at the edge of their zone of proximal development, which means that they could not generate and utilize the experiment strategies on the basis of their own reasoning capabilities, but could accomplish it by resorting to the provided experimental support. Therefore, ES could have acted as the more positive treatment than the negative one for these learners. Learners with high reasoning ability could avoid the negative influence of ES because they had fairly stable ability to deal with the discovery tasks. The negative influence of the extra cognitive load caused by ES could have more prominent interference among the middle-band learners. If this is the case, learning support in a simulation environment must be adapted to the levels of the learners to maximize the benefits of the support and avoid the possible negative influence due to extra cognitive load. Further research needs to explore the possible cognitive load caused by the experimental support and to examine the effect of adaptive experimental support in simulation-based discovery learning.

ES also manifested a notable negative main effect on the intuitive understanding test. One possible reason lies in the inconsistency between the focusing feature of ES and the variations of the problem situations in the intuitive understanding test. The experimental design guidance included in ES highlighted the importance of 'varying one thing at a time' in conducting experiments, whereas the intuitive understanding test provided learners with the pairs of instances involving more than one factor varied. Learners were required to predict the comparison of the sizes of the upthrusts acting upon the pairs of objects that might differ on more than one variable. The feature discrepancy between ES and the test might have led to negative transfer.

RS, designed to promote learners' reflective generalization of discovery experiences, encompassed such treatments as tracing and displaying learners' discovery processes in detail, guiding them to reflect over each experiment, and prompting learners to reflect over all the experiments and conclusions at the end of discovery. In the results, RS demonstrated a notable main effect on the post-tests of knowledge integration and variable control skills. Learners receiving RS had formed stronger and clearer associations between the discovered rules of upthrust and such core background concepts as 'balance of forces'. Also, RS group learners had grasped the strategies of controlled experiment more effectively through the discovery activities. RS could raise learners' self-awareness of the exploration processes and promote them to review and reflect over how they had examined the effect of shape, volume, and mass on the size of upthrust. This was helpful for them to develop integrative conceptual understandings and efficient procedural skills. In a study examining the role of reflection in learning variable control strategies, Lin and Lehman (1999) found that it was meaningful to prompt learners to reflect upon and justify their experimental activities in order to help learners to acquire efficient variable control skills. This conclusion is consistent with the outcome of our present study. So far, research on the self-regulation of discovery learning has been frequently focused on the self-monitoring of the valid design of experiments and the systematic manipulations and observations of variables (e.g. Schauble *et al*. 1993); rarely have they addressed the issues of reflective abstraction and generalization. This study was aimed to make progress in this direction. Similarly, the importance of reflective generalization has also been stressed in the research of other forms of explorative learning such as problem-based learning, stating that reflective generalization is an important phase for learners to construct elaborate, integrative, and transferable knowledge and skills through problem solving activities (Hmelo & Ferrari 1997).

Conclusion

By examining the effects of ES, IS, and RS on simulation-based SDL, this research inquiry has indicated and discussed how the generative meaning-making process for problem representation and hypothesis generation, the systematic and logical scientific experiment process, and the reflective generalization of discovery experiences constitute three critical interrelated perspectives in SDL. Learning supports, either embedded within simulation software or provided by human tutors in classroom settings, should be directed towards all three perspectives to invite meaningful, systematic, and reflective discovery learning with computer simulations.

In future research, more repetitive examinations of the three types of learning support (especially ES and RS) need to be conducted. Also, the research of simulation-based SDL so far has been based on a laboratory-focused view of scientific practice and overlooked the social discourse process in knowledge building communities (Scardamalia, 2002). The discovery tasks are often straightforward and rule-based problems that do not require deepening explanations. The inquiry processes and tools are mostly prespecified, leaving little space for emergent activities. Therefore, future research in this field needs to integrate scientific discovery learning into knowledge building communities (Chen & Zhang, 2002) and capture the social and dynamic nature of authentic inquiry activities.

Acknowledgements

This work was partially supported by the National Educational Science Plan Funding (EMB010888). We are grateful to Dr Cameron Richards from the Hong Kong Institute of Education for valuable comments and criticism on the earlier version of this article.

References

Ausubel D.P., Novak J.D. & Hanesian H. (1978) *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston, Inc., New York.

Chen Q. & Zhang J. (May 2002) Collaborative discovery learning based on computer simulation. Paper presented on the *18th Annual Rutgers Invitational Symposium on Education (RISE): Using Technology to Enhance Learning*, Rutgers, NJ.

Chi M.T.H., Feltovich P.J. & Glaser R. (1981) Categorization and representation of physics problems by experts and novices. *Cognitive Science* **11**, 121–152.

de Jong T. & van Joolingen W.R. (1998) Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research* **68**, 179–201.

de Jong T., Martin E., Zamarro J., Esquembre F., Swaak J. & van Joolingen W.R. (1999) The integration of computer simulation and learning support: an example from physics domain of collisions. *Journal of Research in Science Teaching* **36**, 597–615.

Glaser R., Schauble L., Raghavan K. & Zeitz C. (1992) Scientific reasoning across different domains. In *Computer-Based Learning Environments and Problem Solving* (eds E. de Corte, M.C. Linn, H. Mandle & L. Verschaffel), pp. 345–373. Springer-Verlag, Berlin.

Hmelo C.E. & Day R. (1999) Contextualized questioning to scaffold learning with simulations. *Computers & Education* **32**, 151–164.

Hmelo C.E. & Ferrari . (1997) The problem-based learning tutorial: cultivating higher order thinking skills. *Journal of the Education of the Gifted* **20**, 401–422.

Klahr D. & Dunbar K. (1988) Dual space search during scientific reasoning. *Cognitive Science* **12**, 1–48.

Kuhn D., Schauble L. & Garcia-Mila M. (1992) Cross-domain development of scientific reasoning. *Cognition and Instruction* **9**, 285–327.

Lee J. (1999) Effectiveness of computer-based instructional simulation: a meta analysis. *International Journal of Instructional Media* **26**, 71–85.

Leutner D. (1993) Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional supports. *Learning and Instruction* **3**, 113–132.

Lewis E.L., Stern J.L. & Linn M.C. (1993) The effect of computer simulations on introductory thermo-dynamics understanding. *Educational Techology* **33**, 45–58.

Lin X. & Lehman J. (1999) Supporting learning of variable control in a computer-based biology environment: effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching* **36**, 837–858.

Njoo M. & de Jong T. (1993) Exploratory learning with a computer simulation for control theory: learning processes and instructional support. *Journal of Research in Science Teaching* **30**, 821–844.

Quinn J. & Alessi S. (1994) The effects of simulation complexity and hypothesis generation strategy on learning. *Journal of Research on Computing in Education* **27**, 75–91.

Reid D.J., Zhang J. & Chen Q. (2003) Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning* **19**, 9–20.

Rivers R.H. & Vockell E. (1987) Computer simulations to simulate scientific problem solving. *Journal of Research in Science Teaching* **24**, 403–415.

Ross J.A. (1988) Controlling variables: a meta-analysis of training studies. *Review of Educational Research* **58**, 405–437.

Ruiz-Primo M.A. & Shavelson R.J. (1996) Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, **33**, 569–600.

Scardamalia M. (2002) Collective cognitive responsibility for the advancement of knowledge. In *Liberal Education in a Knowledge Society* (ed. B. Smith), pp. 67–98. Open Court, Chicago.

Schauble L., Raghavan K. & Glaser R. (1993) The discovery and reflection notation: a graphic trace for supporting self regulation in computer-based laboratories. In *Computers as Cognitive Tools* (eds S.P. Lajoie & S.J. Derry), pp. 319–337. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Shute V.J. & Glaser R. (1990) A large scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments* **1**, 51–77.

Swaak J. & de Jong T. (2001) Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning* **17**, 284–294.

Swaak J., van Joolingen W.R., Wouter R. & de Jong T. (1998) Support for simulation-based learning: the effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction* **5**, 235–252.

Tabak I., Smith B.K., Sandoval W.A. & Reister B.J. (1996) Combining general and domain-specific strategic supports for biological inquiry. In *Intelligent Tutoring Systems* (eds C. Frasson, G. Gauthier & A. Lesgold), pp. 288–297. Springer-Verlag, Berlin.

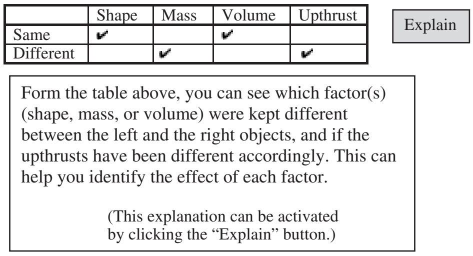
van Joolingen W.R. & de Jong T. (1997) An extended dual search space model of scientific discovery learning. *Instructional Science* **25**, 307–346.

Veenman M.V.J. & Elshout J.J. (1995) Differential effects of instructional support on learning in simulation environments. *Instructional Science* **22**, 363–383.

Zhang J. (2000) *Simulation-based scientific discovery learning: the internal conditions and the facilitative learning supports*. PhD Dissertation, submitted to Beijing Normal University, China.

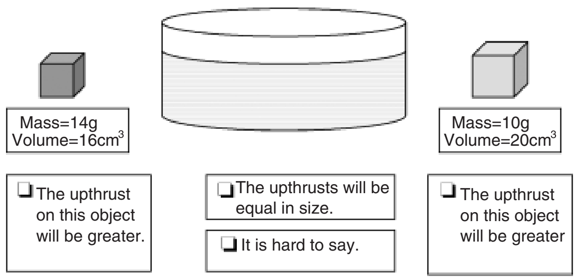
Appendix I

An example showing which variable(s) were the same or different in two chosen objects is given below.



Appendix II

An example of the items assessing intuitive understanding is as follows. One object has a greater mass but a smaller volume than the other. They both keep floating in water. How will the size of the upthrust on each object compare?



Appendix III

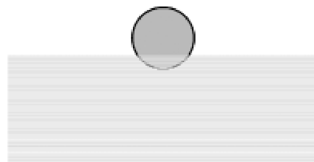
Focused examination of volume: an example that was given full score.

No.	Shape	Mass (g)	Volume (cm^3)	Upthrust (N)
1 left	Ball	18.00	30.00	0.18
1 right	Ball	18.00	20.00	0.18
…				
3 left	Box	5.00	10.00	0.05
3 right	Box	5.00	400.00	0.05

In this case, the first and third pairs of experiments are both focused on the volume, which is the only input variable that was kept different between the left and right sides. Also, these two pairs of experiments examined the effect of volume at different levels of the mass (18.00 and 5.00 g) and shape (ball and box).

Appendix IV

Examples of the 'instance-clustering' items in the knowledge integration test are given below.



Typical instance: A beach ball floating in a swimming pool.

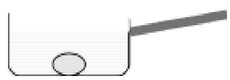
Please tick the 'YES' or 'NO' box for each of the following instances (a–d) to show if it is similar to this instance (the floating beach ball), and explain why you think so.

(a)

A book lying on the desk.

This instance differs from the typical instance in surface features but involves the same force pattern.
This instance is more similar to the typical instance in surface features but involves a different force pattern.

(b)



An egg in a saucepan half full of water.