Error-Based Simulation for Error-Awareness in Learning Mechanics: An Evaluation

Tomoya Horiguchi1*, Isao Imai2, Takahito Toumoto3 and Tsukasa Hirashima4

1Graduate School of Maritime Sciences, Kobe University, 5-1-1, Fukaeiminami, Higashinada, Kobe, Hyogo 658-0022, Japan // 2Hanazono Junior High School, 4-1-1, Hanazono, Hanamigawa, Chiba 262-0025, Japan // 3Faculty of Engineering Division II, Tokyo University of Science, 1-3, Kagurazaka, Shinjyuku, Tokyo 162-8601, Japan // 4Department of Information Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashihiroshima, Hiroshima 739-8527, Japan // horiguti@maritime.kobe-u.ac.jp // isao-1@mx3.mesh.ne.jp // tomoto@ms.kagu.tus.ac.jp // tsukasa@iel.hiroshima-u.ac.jp

*Corresponding author

(Submitted October 23, 2012; Revised July 9, 2013; Accepted October 24, 2013)

ABSTRACT

Error-based simulation (EBS) has been developed to generate phenomena by using students’ erroneous ideas and also offers promise for promoting students’ awareness of errors. In this paper, we report the evaluation of EBS used in learning “normal reaction” in a junior high school. An EBS class, where students learned the concept with EBS, was compared with a usual class where students learned it as usual. We used a pre-test, post-test, delayed post-test, and an interview after the delayed post-test. In tests, three kinds of tasks were included. A learning task was used in the learning process. A complex task is a little different from a learning task and can’t be solved by memorization. A transfer task is very different from the learning task and can’t be solved without generalizing the learning. In the post-test and delayed post-test, the scores of the EBS class were significantly higher than those of the usual class. The differences of scores between these classes in the complex and transfer tasks were larger than those in the learning task. In the interview, students in the EBS class explained their solution in more conceptual ways. These results suggest that EBS contributed to the generalization and retention of the learned concept, “normal reaction.”

Keywords

Learning from errors, Awareness, Misconception, Error-based simulation

Introduction

Natural science explains and predicts natural phenomena. One of the most important purposes of elementary science education is to make students able to explain and predict natural phenomena with scientific concepts. It is, however, often difficult for students to connect abstract scientific concepts to concrete natural phenomena. As well, disconnection of them causes several serious misconceptions (Driver, Guesne, & Tiberghien, 1985; Osborne & Freyberg, 1985). Therefore, supporting students’ comprehension of this connection is a very important issue in elementary science education.

Scientific experiment or demonstration is a popular teaching method to connect them. First, a phenomenon is shown to the students, and then it is explained with scientific concepts that are the targets of teaching. Simulation-based learning environments (SLE) have been investigated to assist such “learning from experiments” and have been found useful for the introduction or acquisition of scientific concepts (Towne, de Jong, & Spada, 1993). When a student makes an error in prediction, showing the correct simulation could be useful for correcting that error. The difference between the correct simulation generated by the SLE and the phenomenon predicted by the student makes the student aware of the error. However, students sometimes have wrong concepts for explaining correct phenomena. In such cases, phenomena generated by SLE aren’t useful because the phenomena are the same as those they connected to the wrong concepts. For example, in elementary mechanics, students often answer that gravity is the only force acting on a block on a table while they predict that the block will stay at rest on the table.

Error-based simulation (EBS) addresses this problem by generating a phenomenon using students’ erroneous ideas to help them become aware of errors when they know some correct phenomena connected to their wrong concepts (Hirashima, Horiguchi, Kashihara, & Toyoda, 1998). In the above example, EBS generates an unnatural phenomenon where the block sinks into the table because gravity, which isn’t cancelled by any other force, causes a downward motion of the block. EBS is a generally useful method to generate counterexamples to students’
misconceptions or erroneous answers, and we have developed several prototype systems not only for mechanics but also for drawing (Matsuda et al., 2003) and English composition (Kunichika, Hirashima, & Takeuchi, 2006).

EBS has an important feature that other systems don’t have that can bring out the pedagogical merit in simulation-based learning. Even when a model is incalculable (e.g., because the constraints specified by the student contradict each other), our approach involves creating an EBS by relaxing some of the student’s model constraint(s). In the above example, the constraint “rigid objects (i.e., the block and table) never overlap” is violated in the simulation. Thus, EBS works as a counterexample, demonstrating that “if students’ ideas were correct, the fundamental constraint in the real world would be violated.” Because of this feature, shown the next section in detail, one can design a learning environment with EBS that has the advantages of discovery learning (Crews, Biswas, Goldman, & Bransford, 1997; Loh, Reiser, Radinsky, Edelson, Gomez, & Marshall, 2000) and directed learning (Klahr, 2009).

So far, preliminary experiments in which small number of university students learned with the above systems and teachers evaluated the functions of these systems suggested the possibility of EBS (Kunichika et al., 2006; Matsuda et al., 2003). Additionally, a preliminary test in which small number of junior high school students learned elementary mechanics with EBS suggested that EBS promoted their conceptual understanding (Horiguchi, Imai, Toumoto, & Hirashima, 2007). Though the test was conducted in a practical situation, the result wasn’t statistically validated because no control group was made for comparison. In this paper, therefore, we describe a practical use and evaluation of EBS in a controlled experiment is described. We used EBS for learning “normal reaction” in a junior high school. The results suggest that EBS contributed to generalization and retention of the learned concept “normal reaction.”

In this paper, we first introduce the outline of EBS, then describe the features of EBS compared with the related work. The purpose of this practice and the procedure of the experiments are described next, followed by its results and the discussion about them. We conclude this paper with some remarks.

**Error-based simulation as a method to promote awareness of errors**

**Error-based simulation**

In order to make students correct their erroneous idea, cognitive conflict often plays an important role (Glynn, Yeany, & Britton, 1991; Osborne & Freyberg, 1985). When the difference between an erroneous answer and the correct one is significant for them, it could cause cognitive conflict. Unexpected correct phenomena shown in experiments or simulation, for example, often have such an impact. An explanation that connects the correct concepts to the phenomena could prompt students to correct their error.

However, when the difference isn’t significant for students, it hardly causes cognitive conflict (Chinn & Brewer, 1993). For example, when a student knows the correct phenomenon and explains it with erroneous concepts, the explanation with the correct concepts may fail to prompt him/her to correct the error because the difference between the erroneous and correct concepts has less impact. In such a case, therefore, the difference should be made more “visible.” That is, what the difference implicates should be clear for students to understand its significance. We call this “error visualization” (Hirashima et al., 1998; Horiguchi et al., 1999).

Figure 1 shows the framework of error visualization with EBS. EBS is generated by mapping errors in symbolic expression to erroneous behavior. In this framework, a student’s answer should be regarded as a model of the target in question from which the behavior is generated. The difference in behavior is better to make students aware of the errors and motivate them to correct the errors. In order to use EBS effectively, we investigated the following three factors: visibility, reliability, and suggestiveness. Visibility concerns whether the difference between normal behavior and EBS is enough to make students aware of the error. Reliability pertains to whether the mapping from symbolic expression to behavior is reliable for students. Suggestiveness pertains to whether the difference between normal behavior and EBS suggests the way to correct the error. These factors are very interesting and important, especially to extend the target domain of EBS (Horiguchi et al., 1999).
We show some examples of EBS with the problem in Figure 2(a). Blocks 2 and 3 are connected with a string through a pulley. The mass of Block 1 is $M$, Block 2 $M$, and Block 3 $2M$. The mass of the string and the pulley is negligible. The acceleration of gravity is $g$. All blocks move without friction. $T$ is the tension of the string between Block 2 and Block 3. Block 2 is restricted to movement along the right side of Block 1. Therefore, normal force, $N$, works between the two blocks, and they move as one towards the left. Because Block 1 moves, relative acceleration should be used in the equation of motion of Block 3. The horizontal component of acceleration of Block 1 is $a_1$, Block 2 $a_2$, and Block 3 $a_3-a_1$. The vertical component of acceleration of Block 2 is $b_2$. Note that $a_2$ is equal to $a_1$, and $a_3$ is equal to $b_2$, respectively. The correct set of equations of this system is shown in Figure 2(b).

Suppose a student set up an erroneous equation $3' \ - M b_2 = Mg - T$ instead of the correct equation, (3). The set of equations $(1), (2), (3'), (4), (5)$ and $(6)$ is incalculable because of over-constraints. In such a case, a constraint (i.e., equation) is deleted to generate simulation. Since EBS should imply the student’s error, equation $3'$ is out of the candidates for deletion. From the viewpoint of visibility, a more fundamental constraint in the real world is preferable for deletion. For example, when equation $(5)$ is deleted, such EBS as shown in Figure 2(c) is generated, in which Block 1 and Block 2 overlap (because this equation maintains the relative acceleration between these blocks).
The student would easily recognize something is wrong with her/his equations. For another example, when equation (6) is deleted, such EBS as shown in Figure 2(d) is generated, in which the string between Block 2 and Block 3 shrinks (because this equation maintains the relative acceleration between these blocks). This EBS might have more suggestiveness than the former because it directly shows the error in the vertical acceleration of Block 2 (i.e., b2), which is included in the very error of this student (i.e., equation (3')). Thus, the effectiveness of EBS can be controlled and estimated following the above factors.

A simulator with the facility for such constraint handling is called “Robust Simulator (RSIM).” We have developed some RSIMs that check the consistency of a set of constraints and relax some of them by using heuristics, if necessary (Horiguchi & Hirashima, 2006; Horiguchi, Hirashima, & Forbus, 2012). In this practice, we embedded the facility for generating EBS into the learning environment, which is specialized in normal force in mechanics and always relaxes the rigid-objects-never-overlap constraint (described later).

![Figure 3. Mechanical problems in learning task](image1)

![Figure 4. Screenshots of a drawing and an EBS](image2)

In this practice, “normal reaction" is a learning target. Students are required to answer existing forces on a mechanical situation by drawing arrows of force. If the students make mistakes, EBS is generated by the students’ erroneous arrows. Most of them have enough visibility, reliability, and suggestiveness.

We introduce an example of EBS used in this practice, as shown in Figure 3. Students are shown a mechanical situation and are required to draw all the forces acting on the objects. They may make an erroneous drawing because of some misconceptions, which are regarded as the externalization of their erroneous idea. Based on the drawing, the acceleration of both objects is calculated with Newton’s second law, and their motion is simulated. In Problem 1 of Figure 3, for example, students often draw only the gravity acting on the block without the corresponding normal force as shown in Figure 4(a). In this case, the block sinks into the floor, as shown in Figure 4(b). We expect that such an unnatural phenomenon becomes a useful counterexample to students’ erroneous ideas and contributes to the correction of the errors with high and intrinsic motivation.

**Feature of EBS: From the viewpoint of constraint handling**

Simulating a model based on students’ erroneous ideas isn’t itself a new method. Many learning environments have been developed in which students construct a model and test it by simulation (Bravo, Joolingan, & deJong, 2006;
Bredeweg, Linnebank, Bouwer, & Liem, 2009; Forbus, Carney, Sherin, & Ureel, 2004; Leelawong & Biswas, 2008). However, EBS is different in that it can bring out the pedagogical merit in the method that other systems can’t. That is, even when a model isn’t calculable because of serious conflict between constraints, our framework generates simulation by relaxing some of those constraints. Models by students often have conflicts between constraints whether the constraints are explicitly represented or not. By choosing the basic constraint(s) to be relaxed, EBS works as a counterexample (e.g., “if the model were correct, the block would sink into the floor”). Thus, students could connect the unnatural phenomena in EBS to their erroneous abstract concept.

Other systems, on the other hand, can’t generate simulation when a model isn’t calculable, so they usually give students corrective feedback on the representation of the model, which points out the erroneous parts of the representation (e.g., erroneous nodes or missing links in a concept map) and instructs students how to fix them. Even when a model is calculable, feedback is often given on its representation. However, such feedback doesn’t take advantage of simulation-based learning, that is, to connect abstract concepts to concrete phenomena. In fact, the limitation of such feedback is reported in several empirical studies (Leelawong & Biswas, 2008; Or-Bach & Bredeweg, 2013), and researchers usually try to combine it with other kinds of feedback, such as metacognitive feedback. In summary, EBS not only simulates the erroneous behavior of a model but also makes it as understandable as possible to students. Other systems rely on non-behavioral feedback when the implication of the behavior is hard to interpret.

Another possible way to deal with an incalculable model without losing the pedagogical merit of simulation is to give students another situation and make them model it. The situation should be modeled with the same target concept as the original situation. If the new model is calculable, the simulation can be generated. Otherwise, further situations are tried. However, as we discuss in the next section, using multiple situations for learning a concept might impose a heavy cognitive load on students. More seriously, since the original situation probably remains not understood, students might fail to understand the concept abstractly to form schematized knowledge.

Feature of EBS: From the viewpoint of analogy

It is well-known that most students have great difficulty in understanding “normal force.” For example, even after a teacher explains the concept in class, students often answer “only gravity” when they are asked what forces are applied on a book on a table. In this case, as we indicated in the previous section, usual experiment or demonstration can’t be used as a method to connect the concept to the phenomenon.

Using “bridging analogies” (Clement, 1993) is a very effective method to solve this problem. In this method, the gap between students’ correct belief and their misconception is bridged by a chain of intermediate analogous situations. For example, suppose students misunderstand the situation of a book on a table, which is called the “target.” First, a situation is introduced to them in which a hand is pushing down a spring on a table. Most students understand the spring pushes back up against the hand. This is called the “anchor.” Then, another situation is introduced in which a book is on a flexible board on a table. Students can understand the board pushes up the book because this situation is similar to that of the anchor. In addition, this situation is similar to that of the target. Therefore, students can connect the anchor to the target, to understand that a “normal force” is applied to the book from the table. Such an intermediate situation that shares features with both the anchor and the target is called a “bridging analogy.” It was reported that using bridging analogies in class effectively activated students’ discussion and scientific thinking, through which they understand the concept normal force (Clement, 1993).

The point of using bridging analogies is that, even in cases when usual experiment or demonstration doesn’t work, it explains the target concept (e.g., normal force) in connection with the phenomenon in some situation(s), instead of only explaining the concept itself. However, this method has the following difficulties: One must find an anchor situation that students can understand correctly. One must also find a chain of situations that bridge the gap between the anchor and the target situation. The gap between every pair of adjacent situations should be sufficiently small to be recognized that they are the same from the viewpoint of the target concept (e.g., in both the “hand Pushing-down-on-spring” and the “book-on-flexible-board” situation, students think upward [normal] force works). On the other hand, the gap should be sufficiently large to be a part of the bridge between the anchor and the target situation. This is a trade-off. Therefore, there is always the possibility that a student fails to recognize adjacent situations as the same (e.g., she/he might not accept upward [normal] force works in the “book-on-flexible-board” situation). In fact,
in the lesson designed by Clement (1993), complementary methods are used to compensate for this weakness, such as the guided discussion about the similarities and differences between situations, and the explanation with a microscopic model of normal force. Therefore, the outcome of the lesson should be regarded as the total effect of combined teaching methods including using bridging analogies.

In learning with EBS, on the other hand, the only situation a student must consider is the target situation. Instead of comparing different situations, she/he tries to explain the target by expressing her/his idea about the physical process(es) working in the situation. Any analogous situations aren’t necessary. More importantly, it is easy for a student to compare her/his different trials because the phenomena occur in the same situation. That is, they can be “well-aligned” (Gentner & Markman, 1997; Markman & Gentner, 1993). It is expected that students can easily see their differences that matter (“alignable differences”), which help them regulate the exploration by themselves to find the solution and form a conceptual understanding of the situation. Additionally, because a problem corresponds to a situation, it is relatively easy to design a sequence of problems of which situations are highly aligned (e.g., “a book on a table” and “a book on top of another on a table”). It is also expected that student can easily see their differences and apply their solution of a problem to other problems with appropriate modification, through which they might abstract the solution to form conceptual understanding and schematized knowledge.

Hypothesis

From the discussion in the previous two sections, we claim that using EBS enables one to design a learning environment that has the advantages of both discovery learning (Crews et al., 1997; Loh et al., 2000) and directed learning (Klahr, 2009). That is, since behavioral feedback is instantly given on students’ every trial (even when the model is inconsistent), using EBS promotes discovery learning. Making trials in the same situation makes it easier for students to self-regulate their exploration. At the same time, since each problem can be solved with EBS in a situation (not multiple situations), one can design a sequence of problems of which situations are highly similar and increasingly complex. Such a sequence of problems works as a guide that helps students generalize previous learning to solve a new problem of a (little) more complicated situation. This is directed learning. Therefore, we hypothesize the following: A carefully designed learning environment with EBS, in which behavioral feedback is instantly given on every trial in a problem and a sequence of well-aligned problems is provided, helps a student not only find the solution of a problem but also understand abstract concepts and form schematized knowledge by itself (without complementary methods).

In the following sections, we present the design of our system and the practice and its result with the system to verify this hypothesis.

Procedure of the practice

Purpose and method of the evaluation

The purpose of this practice is to evaluate the effect of EBS by comparing it to the effect of usual teaching from the viewpoints of transfer and retention of learning. To evaluate the transfer, we prepared three kinds of tasks. The first is a learning task that is composed of three problems, shown in Figure 3. They are used not only in the learning phase but also in all tests. Because these problems are used in the class, it is possible to gain a good score only by memorizing the correct answers. The second is a complex task composed of two problems shown in Figure 5, which consist of the same components as the problems in the learning task, but with a different number of components. Therefore, the problems are similar to the learning task, but it is impossible to gain a good score just by memorizing the answers. Generalization of number of components is required to solve the complex task. These problems were used in the post-test and delayed post-test but were not used in the pre-test and learning phase (because the teacher with 26 years of experience who conducted this practice judged the complex task was too difficult for the students as pre-test). The third is a transfer task composed of seven problems. Two problems are exampled in Figure 6. They consist of different components from the problems in the learning task. Therefore, in order to gain a good score, it is necessary to abstractly understand the relation between force and motion, not depending on the components. These problems were also used in the post-test and delayed post-test but not used in the pre-test and learning phase (because of the same reason as the complex task).
As for the examination of retention of the learning effect, we carried out the delayed post-test three months later, in addition to the post-test. We also had an interview with the students to understand how they solved the problems within one day after the day of the delayed post-test.

Learning environment with EBS

For this practice, we used a learning environment that generates EBS based on students’ erroneous solutions in mechanics problems. In learning with the system, a student is provided with three problems of the learning task one by one and required to draw all forces acting on objects. After completing a drawing using a mouse, the student clicks the “done” button to see the behavior of the objects. This is called the learning phase.

In drawings, the points on which forces are acting are specified only in the neighborhood of objects’ centers or edges. The directions of forces can be specified only vertically or horizontally. The magnitudes of forces, that is, the length of arrows, can be selected from large, medium, and small. When the points, directions, and magnitudes of all forces are drawn correctly, natural motion is generated. When there are any mistakes, EBS is generated. A student can modify her/his drawing and see EBSs any number of times, until she/he completes the correct drawing for the current problem.

Occasionally, the motion of EBS is similar to natural motion. For example, when no forces are drawn in Problem 1, the block stays at rest on the floor correctly. This is the issue of visibility. In all problems used in this practice, the natural behavior is motionless. When an EBS doesn’t have enough visibility, the system directly indicates errors in the drawing. The methods to judge the visibility of EBS for moving objects with qualitative reasoning techniques were reported in Hirashima et al. (1998).

Lessons

This experiment was carried out with students who were in the first year of junior high school (grade 7). They were originally divided into three classes, with a total of 84 students. We assigned two of them to EBS class (54 students) and one to usual class (30 students). First, all classes worked on the pre-test. Then, all classes were provided with a lecture as usual in one class time (45 minutes) and only EBS class had additional learning time of another 45 minutes class, during which they solve three problems of learning task with the system. Therefore, the difference between
EBS class and usual class is learning with EBS. In the learning phase, each student used one system with one computer. EBS class worked on the post-test after the learning phase, while usual class after the lecture. Finally, three months later, all classes worked on the delayed post-test and were interviewed by the teacher. The pre-test included only learning task, while the post-test and delayed post-test included all three tasks. All tests were written tests.

All classes were taught by the same teacher who was in charge of science for junior high school students. In the learning phase, one assistant teacher was provided in addition to the class teacher. The teachers helped the students to use the system, while they didn’t give any hints about the solution of the problems. They also carefully observed the students’ activity. Especially, if they observed any student being seriously confused by unnatural phenomena, they were ready to stop her/him.

**Results**

**Students’ learning activities**

In the use of EBS system, all students were actively working on the problems. No student was confused by unnatural phenomena. When students saw unnatural phenomena, we observed that they were motivated to think about the cause of the error in their solutions. No students had any serious difficulties in using the system. All students completed the three problems correctly in the learning phase.

**Results of scores**

The results of the average scores are shown in Figures 7, 8, and 9. They show the effect of the conditions, the effect of the tests and the effect of the tasks, respectively. The statistical analysis is also been summarized in Tables 1, 2, and 3. The marking system for the tests was one point for one correct answer of an acting force; therefore the total mark for the learning task was 14 points, while for the complex task and the transfer task the marks were 19 and 30, respectively.

Figure 7 shows that in post- and delayed post-tests, and in all tasks, the scores of EBS class are higher than those of usual class (while, as shown below, there wasn’t significant difference between EBS and usual classes in pre-test). Especially, greater EBS’s effect was observed in delayed post-tests than in post-tests, and greater EBS’s effect was also observed in transfer tasks than in learning and complex tasks. Figure 8 shows that the score decreases in delayed post-tests were smaller in EBS class than in usual class. Figure 9 shows that the difficulty of tasks increases in the order of learning, complex and transfer task.

![Figure 7. Effect of conditions (Usual vs. EBS)](image)

![Figure 8. Effect of tests (Post vs. Delayed post)](image)
In summary, in post-test and delayed post-test, the scores of EBS class were higher than those of usual class as to all tasks. Especially, more EBS’s effect was observed in delayed post-tests than in post-tests, and so was in transfer tasks than in learning and complex tasks. Note that generalization of the learning result is necessary for answering the problems of complex and transfer task correctly. We can, therefore, conclude that the generalization and retention of the learning results were well done in the EBS class than in the usual class.

**Table 1. Simple-simple main effects of class**

<table>
<thead>
<tr>
<th>Task Factor</th>
<th>Usual class (n = 30)</th>
<th>EBS class (n = 54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning task (full marks = 14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>3.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Post</td>
<td>12.7</td>
<td>13.8</td>
</tr>
<tr>
<td>Delayed</td>
<td>8.6</td>
<td>12.3</td>
</tr>
<tr>
<td>Complex task (19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>16.7</td>
<td>18.4</td>
</tr>
<tr>
<td>Delayed</td>
<td>10.6</td>
<td>16.4</td>
</tr>
<tr>
<td>Transfer task (30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>17.3</td>
<td>22.4</td>
</tr>
<tr>
<td>Delayed</td>
<td>11.5</td>
<td>18.1</td>
</tr>
</tbody>
</table>

```
F = 2.058 p > .10
F = 6.762 p < .01
F = 70.912 p < .001
F = 8.628 p < .005
F = 89.047 p < .001
F = 17.3 p < .001
F = 48.678 p < .001
```

Average scores and SD

**Italic**: small effect (≥ .01 and < .06), underlined: medium effect (≥ .06 and < .14), bold: large effect (≥ .14) (using Cohen’s criteria [Cohen, 1998])

**Table 2. Simple-Simple Main Effects of Task**

<table>
<thead>
<tr>
<th>Task Factor</th>
<th>Usual class</th>
<th>EBS class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>F = 133.212</td>
<td>F = 57.229</td>
</tr>
<tr>
<td>Delayed</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>F = 73.173</td>
<td>F = 93.016</td>
</tr>
<tr>
<td>Delayed</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>F = 95.016</td>
<td>F = 48.678</td>
</tr>
<tr>
<td>Delayed</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>

**Table 3. Simple-Simple Main Effects of Test**

<table>
<thead>
<tr>
<th>Test Factor</th>
<th>Usual class</th>
<th>EBS class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F = 274.796</td>
<td>F = 520.901</td>
<td></td>
</tr>
<tr>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F = 226.940</td>
<td>F = 464.900</td>
<td></td>
</tr>
<tr>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F = 750.349</td>
<td>F = 416.602</td>
<td></td>
</tr>
<tr>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td></td>
</tr>
</tbody>
</table>

Three-factor ANOVA of 2 (class: EBS/usual) x 3 (task: learning/complex/transfer) x 3 (test: pre, post, delayed post) revealed the above observations have statistical significance. Because the interaction of the three factors was significant, and was so for all combinations of two factors, we tested the simple-simple main effect of each factor. Table 1 shows the simple-simple main effects of class factor and their effect sizes. There were significant differences in post- and delayed post-tests ($F = 6.762, p < .01$ and $F = 70.912, p < .001$ in learning task; $F = 8.628, p < .005$ and
$F = 89.047, p < .001$ in complex task; $F = 27.681, p < .001$ and $F = 48.678, p < .001$ in transfer task) while there wasn’t in pre-tests ($F = 2.058, p > .10$). Additionally, the effect sizes were greater in delayed post-tests, and in more difficult tasks ($\eta^2_p = .0273$ for post-test and $\eta^2_p = .281$ for delayed post-test in learning task; $\eta^2_p = .0352$ for post-test and $\eta^2_p = .354$ for delayed post-test in complex task; $\eta^2_p = .109$ for post-test and $\eta^2_p = .193$ for delayed post-test in transfer task). The effect size is calculated by squaring the correlation ratio, and indicated using Cohen’s criteria (small effect: $\geq .1$ and $< .6$, medium effect: $\geq .6$ and $< .14$, large effect: $\geq .14$) (Cohen, 1988). Tables 2 and 3 show the simple-simple main effects of task and test, respectively. According to the former, the differences of difficulty between tasks were significant ($F = 133.212, p < .001$ for post-test and $F = 57.229, p < .001$ for delayed post-test in usual class; $F = 73.173, p < .001$ for post-test and $F = 95.016, p < .001$ for delayed post-test in EBS class). According to the latter, the differences of scores between tests were also significant ($F = 274.796, p < .001$ for learning task, $F = 520.901, p < .001$ for complex task and $F = 226.940, p < .001$ for transfer test in usual class; $F = 464.900, p < .001$ for learning task, $F = 750.349, p < .001$ for complex task and $F = 416.602, p < .001$ for transfer task in EBS class).

**Results of interviews**

Within one day after the delayed post-test, the same teacher who conducted the practical use of EBS interviewed the students. During the interview, teacher only asked them to explain how they solved all seven problems in the transfer task. A total of 42 students from the EBS class and 27 students from the usual class were interviewed. The rest of the students were absent from the interview because it was conducted out of class time. The average scores of the transfer task in the delayed post-test were 18.3 in EBS class and 11.9 in usual class.

The most frequent explanation in usual class was “balance of forces.” If the students explained they tried to balance forces to solve a problem, we categorized their explanations as “force balance.” In usual class, 4.6 problems in 9 problems in average were categorized as “force balance.” Some of the students answered, “they had no idea” or “they found only gravity in a problem” (3.2 problems in average). We categorize these answers as “no idea/only gravity.” No students referred to the motion of objects (Table 4).

**Table 4. Numbers of problems explained with categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Usual class (n = 27, ave = 11.9)</th>
<th>EBS class (n = 42, ave = 18.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>force balance</td>
<td>4.6</td>
<td>1.3</td>
</tr>
<tr>
<td>force and motion</td>
<td>0</td>
<td>4.8</td>
</tr>
<tr>
<td>no idea/only gravity</td>
<td>3.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Table 5. Correlations between delayed post-test scores and numbers of problems explained with categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Correlation</th>
<th>Usual class (n = 27)</th>
<th>EBS class (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>force balance</td>
<td>$r = .84^*$</td>
<td>$p = 2.0E-05$</td>
<td>$r = -.22$</td>
</tr>
<tr>
<td>force and motion</td>
<td>$r = 4.8^*$</td>
<td>$p = .0018$</td>
<td>$r = .44^*$</td>
</tr>
<tr>
<td>no idea/only gravity</td>
<td>$r = -.84^{**}$</td>
<td>$p = 1.9E-05$</td>
<td>$r = .44^{**}$</td>
</tr>
</tbody>
</table>

$^{**}$strong positive correlation, $^*$medium positive correlation, $^{***}$strong negative correlation, $^{**}$medium negative correlation

The relationship between the delayed post-test scores and the number of problems explained with “force balance” was investigated using Spearman rank correlation test (Table 5). There was a strong positive correlation between the two variables, $r_p = .84, p = 2.0E-05$. There was a strong negative correlation between the delayed post-test scores and that with “no idea/only gravity”: $r_p = -.84, p = 1.9E-05$. These results suggest that the students in usual class solved the problems mainly with “force balance” but did not consider the relation between force and motion.
In EBS class, by contrast, the most frequent explanations were “connection between force and motion” (4.8 problems in average). If the students’ explanations suggested that they mentally simulated the effect of forces on objects’ motion, we categorized them as “force and motion” (for example, “Since the gravity was drawing the block downward, I added the upward force for it not to fall”). On the other hand, “force balance” and “no idea/only gravity” was 1.3 and 1.0 problems in average, respectively (Table 4). There was a medium positive correlation between the delayed post-test scores and the number of problems explained with “force and motion,” $r_s = .48, p = .0018$. There was a medium negative correlation between the delayed post-test scores and that with “no idea/only gravity,” $r_s = -.44, p = .004$. The correlation between the delayed post-test scores and “force balance” was not significant, $r_s = -.22, p = .15$ (Table 5). These results suggest that the students in EBS class solved the problems mainly by considering the relations between force and motion.

These results suggest that the awareness for the relation between force and motion plays a crucial role in the difference between usual class students and EBS class students. Though the usual class students who didn’t consider the relation between forces and motions also got better delayed post-test scores compared with pre-test scores, their delayed post-test scores were significantly lower than EBS class students’ in all tasks. Additionally, in all tasks, the decreases between post-test and delayed post-test were smaller in EBS class than in the usual class. As indicated previously, these results suggest that EBS class students got qualitatively better understanding. We suppose it is because they considered the relation between forces and motions. EBS is a method to visualize errors by showing unnatural behaviors of objects based on students’ errors. In this practice, motion of objects was generated by using students’ erroneous forces. Therefore, it is suggested that students not only were aware of errors by observing unnatural behaviors, but also noticed the importance of the connections between force and motion. This can be considered as a normal and also general approach to solve problems in this practice. These results suggest that EBS is useful to promote students to consider the connections between forces acting on objects and their motions, although we have to examine further experiments in mechanics or other learning domains.

Discussion and concluding remarks

We previously hypothesized that a carefully designed learning environment with EBS helps a student to not only find the solution of a problem but also understand it conceptually to form schematized knowledge. The results of our practice strongly suggest this hypothesis is true. Though the teachers didn’t give any hints, all students of EBS class solved all the problems correctly by using the system. Additionally, greater EBS’s effect in delayed post-tests and transfer tasks suggests that the students of EBS class gained a conceptual understanding of which quality was confirmed in the interview with them. (To ensure fairness, we must point out that the EBS class spent more time on learning than usual class, the influence of which should be investigated in future work.)

Note that the system used in this practice doesn’t provide any “intelligent” help. It merely simulates objects’ behavior based on students’ drawing of forces by violating the “rigid-objects-never-overlap” constraint, if necessary. Without corrective feedback on the drawings, students could complete them correctly only by observing the simulation. Without metacognitive feedback, students could generalize and schematize their knowledge for future use. This outcome was, we think, obtained from the design of the system in which (a) instant feedback was given on each trial in the same situation that promoted the alignment of the trials in a problem, and (b) because the situations of a sequence of problems are highly aligned, the alignment of the solutions of the problems is promoted. Additionally, this design became possible because of the features of EBS: (1) The behavior of models can be always simulated even if they aren’t calculable, (2) The unnaturalness of EBS is controlled taking into account its visibility, reliability, and suggestiveness, (3) Students can receive meaningful behavioral feedback on their every trial in a situation (it needn’t be changed), and (4) Such correspondence between a situation and a problem makes it easier to design a well-aligned sequence of problems. Though EBS is on the lines of “modeling and simulation” methodology, its own feature presents an alternative (or a complement) to the existing methods for designing assistance in SLEs. We verified that our method does work for promoting students’ deeper understanding.

Two issues remain to be further investigated: applicability to more complex tasks and metacognitive skills students can acquire. First, the tasks used in this practice were relatively simple because they dealt with mainly one (critical) concept, “normal force.” The interface used for constructing models was also made very simple. Such simplicity might help students concentrate on comparing trials and problems to generalize what they learned. In other learning environments, more complex modeling tasks are dealt with that include a lot of parameters and difficult concepts.
(such as “stock and flow”). It is our important future work to apply EBS to other domains and more complex tasks. Second, since the results of this practice suggest the students of EBS class generalized their knowledge well, we can expect that they acquired some metacognitive skills in reflection. For generalizing and schematizing the knowledge they learned, students might reflect their solutions and problem-solving processes and acquire skills such as how to distinguish between problem-specific and generalizable elements, how to compare problems/solutions, and how to find the essence of a problem/solution. Some of these skills would be domain-independent. Their higher performance in transfer tasks suggests they had got such skills, while they might do well without such skills because our system was designed to strongly guide their learning. The detail is unknown. More elaborate experimentation is necessary to clarify this.

Currently, this is a case study. It is imperative to try EBS in other domains and more complex tasks to investigate the above issues, which would clarify its applicability, (metacognitive) learning effect, and limitation, and how it works in combination with other teaching technologies.

Acknowledgments

The authors would like to thank Prof. Dedre Gentner and Prof. Ken Forbus for their kind advices and important suggestions, which greatly improved this paper. We also thank the members of their research group, the discussion with whom was very fruitful.

References


Forbus, K. D., Carney, K., Sherin, B., & Ureel, L. (2004, August). *Qualitative modeling for middle-school students*. Paper presented at the 18th International Workshop on Qualitative Reasoning (QR04), Evanston, IL, USA.


