Aikawa et al. *Research and Practice in Technology Enhanced Learning* (2024) 19:26
https://doi.org/10.58459/rptel.2024.19026

Research and Practice in
Technology Enhanced Learning

RESEARCH Free and Open Access

Classroom implementation of an auxiliary problem presentation system for mechanics adapted to learners' errors

Nonoka Aikawa [1], Shintaro Maeda [2], Tomohiro Mogi [2], Kento Koike [3], Takahito Tomoto [4] *, Isao Imai [5], Tomoya Horiguchi [6] and Tsukasa Hirashima [7]

*Correspondence:
tomoto@net.it-chiba.ac.jp
Faculty of Information and Computer Science,
Chiba Institute of Technology,
2-17-1 Tsudanuma,
Narashino-shi,
Chiba-ken, 275-0016, Japan
Full list of author information is available at the end of the article

Abstract

Error-based Simulation (EBS) is a learning support framework that visualizes learners' errors and encourages trial and error. However, when a learner is stuck, EBS has difficulty in helping them overcome the impasse. Additionally, giving a correct answer to a learner who is stuck may interfere with the trial-and-error activity that EBS is oriented toward. Therefore, it is necessary to encourage learners during trial-and-error activities without giving them correct answers. In this study, we confirm the effectiveness of our system, which is based on conventional mechanics EBS and provides adaptive auxiliary problems based on learners' errors. Furthermore, we analyze force-based self-overcoming to evaluate our system. Self-overcoming means that the learner can eliminate errors by using the system without the intervention of the teacher. If self-overcoming occurs, the learner can continue trial-and-error with the auxiliary problems, even if they are stuck. To verify the learning effectiveness of such a system, we conducted a classroom implementation with 86 third-year junior high-school learners and analyzed the results. The system logs from the exercises revealed that self-overcoming was taking place, and that it was reflected in the test results.

Keywords: Elementary mechanics, Learning support system, Error-based Simulation, Classroom practice

Introduction

Error-based Simulation (EBS) is an effective learning support framework that encourages trial and error by visualizing learners' errors. EBS first presents a problem and a simulation of what would happen if the learner's (wrong) answer was correct. If the answer is incorrect, the simulation behaves strangely, providing them with an opportunity to recognize their error by observing it. Applications of EBS have been studied in various fields and its

© The Author(s). 2023 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

effectiveness has been demonstrated (Hirashima et al., 1998; Horiguchi et al., 2014; Shinohara et al., 2015).

However, if a learner gets stuck in the middle of the trial-and-error process, it is difficult for EBS to help them overcome the impasse. Mechanics EBS visualizes the motion of an object in response to drawn forces, but it may not be able to visualize complex problems with many forces acting on the object, and thus may not be able to guide the learner to the correct solution. However, providing the correct answer may interfere with the trial-and-error process on which EBS is based, so this should not be done. The goal, then, is for learners to self-overcome, solving incorrect problems without intervention by the teacher.

In this study, we introduced a framework that presents error-specific auxiliary problems to learners who repeatedly make errors on the same problems. This system was then implemented in a classroom implementation for 86 third-year junior high school students (74 of whom constituted the data to be analyzed). We then investigated whether these auxiliary problems enable self-overcoming in learners.

Related work

Mechanics learning support system

Mechanics has been a favored subject in the study of learning support systems, especially since the early days of their development (Wenger, 1987). Andes is a representative example of such a system (VanLehn et al., 2005), providing adaptive feedback by asking the learner to formulate equations and interpreting them to provide hints and explanations related to their errors.

There is some concern that intervention from the system, such as providing hints and explanations, may inhibit learning when an emphasis is placed on recognition of errors (Tulis et al., 2016). Consequently, many simulation-based learning environments that provide feedback to a learner's input have been proposed, and their effectiveness has been demonstrated for mechanics (Park, 2019).

Error-based Simulation (EBS)

Error visualization is a framework that makes learners aware of their errors by visualizing them in the target learning task, rather than giving it to them directly. A learning environment using simulation, especially error visualization, is referred to as an EBS (Hirashima et al., 1998).

Applying EBS to mechanics has been shown to promote a greater conceptual understanding. This understanding helps to increase learning retention (Wilcox et al., 2020). These benefits have been observed to persist for one to three months (Horiguchi et al., 2014;

Shinohara et al., 2015). In the present study, we deal specifically with EBS for mechanics (mechanics EBS).

It has also been pointed out that mechanics EBS may not allow learners to properly correct errors when visualized. Imai et al. (2008) conducted an EBS class practice session at a junior high school. While the learning effects of EBS were confirmed, the authors noted that EBS has limitations in its effectiveness for phenomena involving complex interactions. Therefore, there are cases in which mechanics EBS is ineffective at guiding learners to a correct solution.

Learners may then get stuck in the problem practice in EBS, which could cause them to lose motivation and give up learning (Beck & Gong, 2013). It is common to give a stuck learner a completely correct answer or a hint about the incorrect portion of their answer (Shute, 2008). However, giving a learner the correct answer may leave them without a deep understanding of the error, and may interfere with the trial-and-error process on which EBS is based, so this should be discouraged.

Scaffolding in learning support systems

Scaffolding is a teacher-supported process that enables learners to perform tasks they cannot complete independently (Wood et al., 1976), and meta-analyses have shown that it improves individual academic performance (Belland et al. 2017).

EBS and modeling learning environments are similar in that they generate and support strange simulations from learners' answers. VanLehn et al. (2016) summarized scaffold methods in modeling learning environments and developed Dragoon, a system that implemented some of these methods. The scaffold method used in this study, and not implemented in Dragoon, is decomposition into subsystems. Beek and Bredeweg (2012) proposed a system that provides a causal explanation as a scaffold for beginning learners to identify the cause of the difference between correct and incorrect behavior in a modeling learning environment. However, providing explanations may lead to passive learning, in which the learner simply reads and accepts the explanations.

Hayashi et al. (2014) developed a physics learning system in which an auxiliary problem that simplifies the original problem is presented to the erring learner. They confirmed that this system facilitates self-overcoming. However, in their system, learners are required to work on auxiliary problems in a predefined order, regardless of the type of error they make.

Burton et al. (1984) discussed increasingly complex microworlds (ICM). ICM is an approach in which learners begin by experiencing and understanding simple phenomena before gradually increasing their complexity, thereby facilitating greater understanding of more complex phenomena. In this study, we instead present the target problem first. The learner's errors are then identified, and simple error-specific auxiliary problems are presented.

Impasse-breaking support in EBS

Aikawa et al. (2020) conducted research on providing adaptive auxiliary problems in mechanics EBS. They developed a system that targets complex problems and the associated impasse and presents auxiliary problems to learners who are stuck. Aikawa et al. used the dynamic whole-task selection approach (Salden et al., 2006), which automatically diagnoses the learner's errors and presents auxiliary problems accordingly. This approach achieves a high level of problem control in which the most complex problem is solved first. The learner is then given a task that allows them to learn partial solutions dynamically based on the problem-solving situation. One way of creating auxiliary problems is to simplify the original problem (Hayashi et al., 2014). Having learners solve these auxiliary problems helps them understand where they are having difficulties in solving the original problem. They then assessed the learning effects for university students. However, some unresolved issues remain in the research of Aikawa et al.:

1. There was no analysis of whether the auxiliary problem helps to resolve the impasse.
2. The participants in their experiment were not junior high school students (beginning mechanics students), the primary target of the system.
3. The sample size was small.

Therefore, in the present study, the number of self-overcoming instances is analyzed for 1. Additionally, we introduce a system that provides auxiliary problems to learners who are stuck in 2 and 3, specifically, to a class practice session for 86 third-year junior high school students (74 of whom were included in the analyzed data), and we verified its learning effects.

In this study, an impasse is defined as "a state in which no progress is made in problem solving due to repeated failures over a long time through trial and error." If self-overcoming occurs without teacher intervention in a learner who repeatedly gave incorrect answers to a problem, they will have broken the impasse by using an auxiliary problem. Based on this, Aikawa et al. analyzed self-overcoming (referred to as "breaking an impasse" in their paper). However, they were able to analyze only a single error in the original problem and up to its solution, and a comprehensive analysis of what errors and self-overcoming occurred during the learning activity was not possible. Here, we calculate the numbers of errors and instances of self-overcoming for each force in the learners' answers, and assess for which ones the auxiliary problems were effective.

System used

This section is an overview of the EBS system. The system works by analyzing the learner's error based on their answer history for the exercises and presenting them with appropriate auxiliary problems.

Specifically, learners first work through exercises in a system similar to that used in conventional EBS (Figures 1 and 2). After a certain number of incorrect attempts at the same problem, the system offers the learner the choice to work on an auxiliary problem, and if they choose to do so, the system automatically presents one to them (Figure 3). The system analyzes the learner's answer history and the errors in their answers, then selects auxiliary problems appropriate for those errors. The learner works on the auxiliary problems, and if they fail a certain number of times, the system presents additional auxiliary problems in the same way. If the learner answers a given auxiliary problem correctly, then the system presents a series of incrementally more complex problems. The goal is for the learner to be able to solve the initial problem by repeating this process.

For this system to work effectively, it is necessary to have a set of auxiliary problems appropriate for the treatment of each error and a problem graph organized based on the differences between problems. The system used in this study implements these features in a conventional EBS, and was itself implemented on a tablet. This allowed us to make adjustments that made it easier for junior high school students to input data, based on the opinions of those who had experience teaching the system. For clarity, we refer to the auxiliary problems in the system as "support problems."

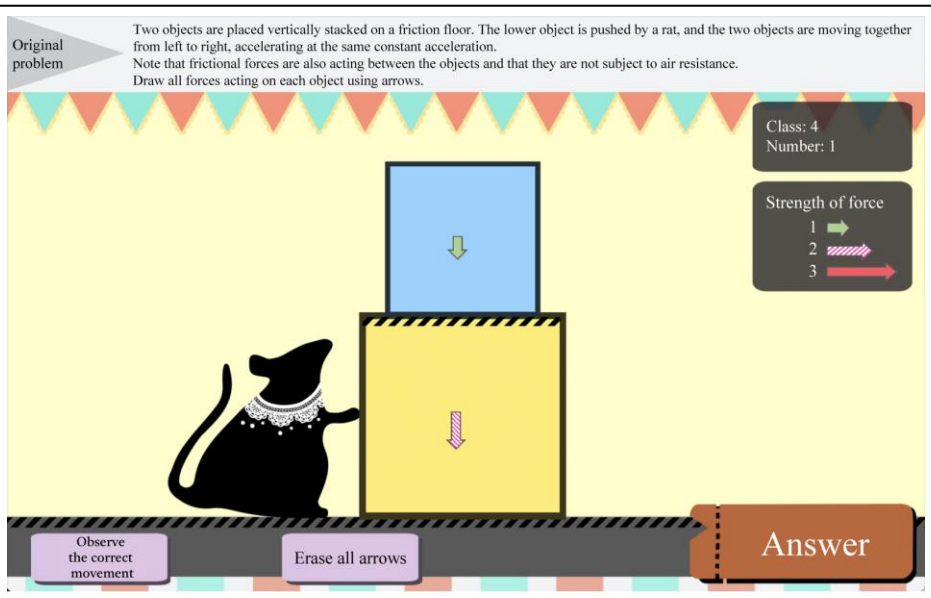


Fig. 1 System screen presenting the original problem

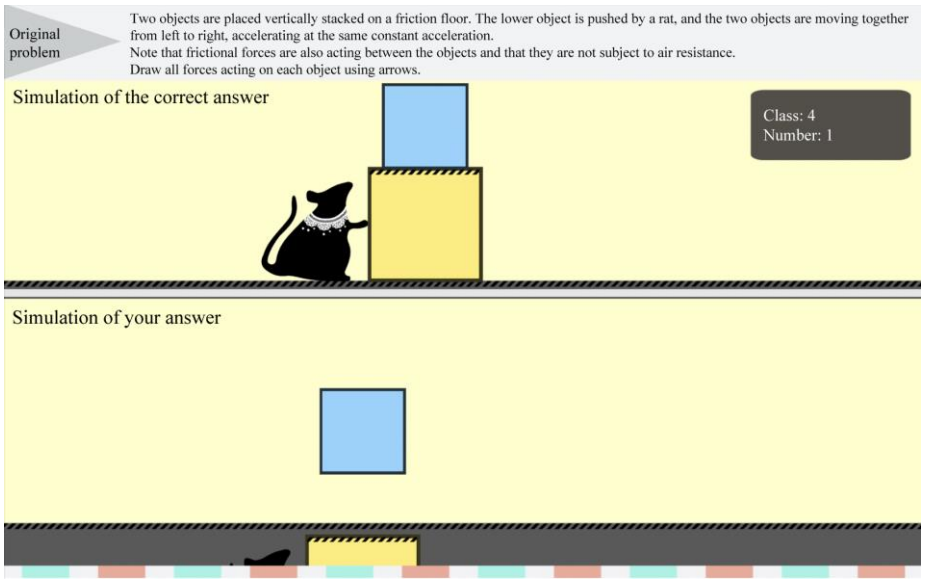


Fig. 2 System screen for simulation

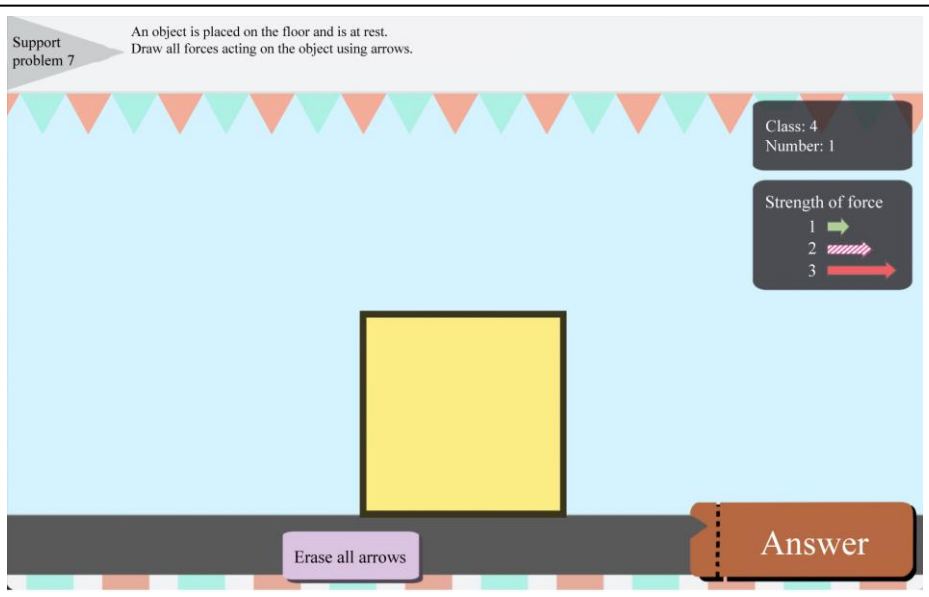


Fig. 3 System screen for presenting support problems

Support problems

Here, we detail the auxiliary problems used by the system. Figure 4 shows all the problems implemented in the system and their problem graph.

The original problem is shown in Figure 4. In this problem, two objects are stacked on a floor with the lower one being pushed sideways by an external force, causing it to

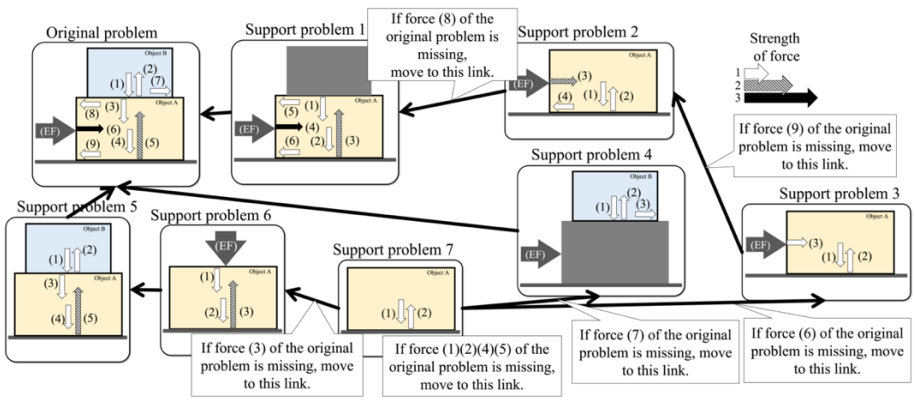


Fig. 4 The problem graph implemented in the system

accelerate. There are nine forces at work here, and we created auxiliary problems for each of them specifically.

For example, support problem 6 includes force (3) (which exerted on the lower object by the upper object) but removes the lateral force and the upper object. This problem replaces the object above with an external force, bringing force (3) to the attention of the learner.

We also included support problem 5 in the problem graph so that learners who have learned force (3) in support problem 6 can apply their understanding to the two-object problem. Support problem 5 is a stationary problem with two overlapping objects, so by solving support problem 6 and then support problem 5, the learner will be able to apply force (3) to problems with two overlapping objects.

The gravitational and normal forces, represented by (1), (2), (4), and (5) in the original problem, are common to all problems, so we used the simplest problem, support problem 7, as an auxiliary problem. For (7), (8), and (9) in the original problem, we created support problems 4, 1, and 2, respectively. This system implemented a problem graph consisting of eight problems: seven support problems and the original problem.

Error analysis

Here, we describe the method used to analyze learners' errors in this system. Recall that we define an impasse as a state in which no progress is made in solving a problem because the learner repeats the same mistake multiple times. The system analyzes the number of errors by counting the ones that they made most frequently based on their answer history.

First, the learner answers a problem by drawing a diagram using arrows. After each answer, the system compares their drawing to the correct answer and counts the number of arrows that are missing.

The system performs this kind of analysis for each answer the learner gives and extracts the force that is missing most often from their answer history. It then assigns this as the

learner's error (the process in the case of a tie is described in the "Presentation of auxiliary problems" subsection). For example, if the learner answers the original problem five times and misses force (3) four times, force (8) three times, and force (5) once, the system will mark force (3), the most frequently missed force, as the error.

Presentation of support problems

Here, we describe how the system presents support problems. It compares the learner's errors using the problem graph in Figure 4 and selects the corresponding link. It then presents support problems in decreasing order of the number of elements at both ends of the link.

For example, if the learner continues to make errors with force (3) in the original problem, then the system will select the corresponding link between support problem 7 and support problem 6. Support problem 7, which has fewer elements, is presented first, and once answered correctly, support problem 6 is then presented. We assume that by solving these problems in this sequence, the learner will become aware of how they differ, which is in the presence of force (3).

If multiple arrows are tied for the most times being excluded from the answer, then the system sets a priority for each arrow and presents the support problems in order of highest priority. In this system, forces (1), (2), (4), and (5) (gravity and the normal forces) have the highest priority, followed by force (6) (an applied external force), force (3) (the force exerted by the lower object on the upper object), and forces (9), (7), and (8) (all friction forces), in that order.

The learner works through these support problems, and after a certain number of incorrect answers, the errors are analyzed again in the manner outlined in the "Analysis of error" subsection, and a new set of support problems are presented. If they answer a support problem correctly, a support problem with more elements is presented according to the problem graph in Figure 4. Since the support problems are constructed from the original problem, they are designed to revert to the original problem as they become more complex, the goal being to help the learner recognize what was prohibiting them from answering it correctly.

As for when support problems are presented, the system is designed to present one after 10 consecutive incorrect answers to the same problem. In consideration of the learner's workload, we also implemented a function that allows the system to present a support problem at the learner's discretion if they make three or more errors in the same problem.

Comparison with the system of Aikawa et al. (2020)

Here, we describe the differences between our system and that developed by Aikawa et al. (2020), specifically regarding the interface, problem graph, and transition method. In

Aikawa et al.'s system, the interface was drawn by inputting certain elements, such as the starting point, direction, and size. However, it is considered to be too much work for junior high school students to have them think of each element. Our system is designed so that they can draw arrows on a tablet using their fingers. The tasks in the problem graph are the same, but support problems 1 and 4 are different. In Aikawa et al.'s system, these problems dealt with the acceleration of an object between the floor and the ceiling and the acceleration of an object on a conveyor belt. The force elements for these are identical to those in the original problem, but the objects are different. However, to facilitate the transition to the original problem, both problems were treated as phenomena in which two objects accelerate when the object below is pushed, thus bringing the objects involved into a sort of unification with those in the original problem. Aikawa et al.'s transition method differed, however, in that it presented a problem that included the force associated with the error when an impasse was reached. However, we considered it necessary to have the participants observe the difference between the problem without the corresponding force and the problem with the corresponding force when there was an error. Therefore, our system presents problems that do not include the force associated with the error before presenting those that do.

Practical use

Procedure

We conducted a class practice session at a junior high school to assess the effectiveness of our system for junior high school students learning elementary mechanics. EBS, which forms the basis of this system, has been practiced many times in junior high schools (Horiguchi et al., 2014; Shinohara et al., 2015). In a study by Horiguchi et al., it was mentioned that none of the junior high school students who used EBS experienced any major difficulties in using the system. In this study, we targeted students who had already studied the range of physics covered by the system as part of their curriculum; the system is not intended for students who do not know the subject matter, but those who do and get stuck in the exercises.

Therefore, three classes (two with 29 students and one with 28 students, for a total of 86) in the third grade of a junior high school were taught for three periods (135 minutes). The class practice session consisted of a greeting and class explanation (41 minutes), a pre-test (10 minutes), a period using the system (64 minutes), a post-test (10 minutes), and a questionnaire (10 minutes). To assess whether the students had retained what they had learned, a delayed test was administered eight weeks later, when they were considered to

have sufficiently forgotten the test problems. Similar studies have allowed a period of one to three months.

Concerning instructional intervention during the use of the system, we implemented a practice mode using simple problems to demonstrate how to use the system and guided students while letting them operate the system. They were then asked to switch to a version of the problem targeted in this classroom implementation. During the use of the system, teachers and system developers visited the students, answered questions about the system, and gave guidance to students who seemed to be stuck. They answered questions about the system specifications but did not give stuck students the answers, instead advising them to "try the answers you think" and to "interpret the simulations."

Evaluation method

Learning effectiveness was evaluated based on the results of the pre-, post-, and delayed tests. The difference between the pre- and post-tests was used to investigate how well the system helps learners solve problems. The difference between these and the delayed test was used to investigate whether the system provides learners with an understanding that persists over time. The content of each test was the same, consisting of four learning tasks (Figure 5) and three transfer tasks (Figure 6), for a total of seven problems. The problems

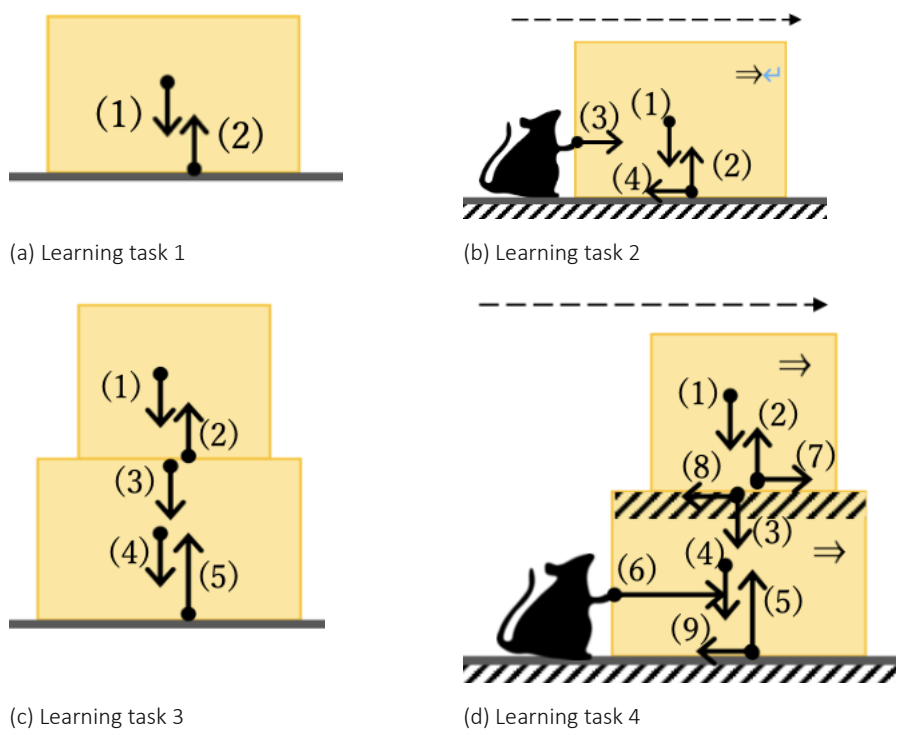


Fig. 5 Learning tasks used in the test

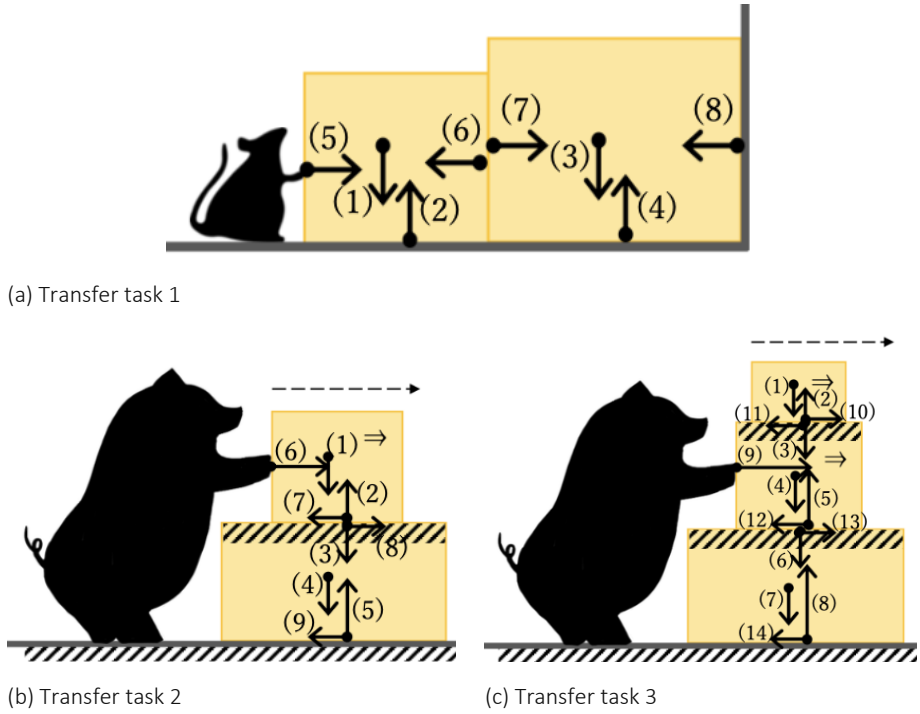


Fig. 6 Transfer tasks used in the test

in the learning tasks were part of those handled by the system: the original problem of Figure 4 and support problems 2, 5, and 7. The problems in the transfer tasks were of a more advanced nature and were not included in the system. All these problems were drafting problems, and the correct answers are indicated with arrows in the diagrams in Figures 5 and 6.

Furthermore, we used the system logs from the exercises to assess whether the auxiliary problems helped to eliminate errors in the learning activities. Specifically, we analyzed the students' answer histories for the original problem before and after solving the auxiliary problems. We then determined whether error resolution occurred in their answers for each ability (we refer to this as self-overcoming (Hayashi et al., 2014) in this paper) and calculated the number of times this occurred.

Results

Test results

The results of the pre-, post-, and delayed tests are given in Table 1. Since some of the students were absent from class and the delayed test, only 74 could be included in the data. In Table 1, we summarize the means and standard deviations for the learning tasks, the

Table 1 Test results. Values outside and inside the parentheses are mean values and standard deviations, respectively. The learning task is worth 4 points, the transfer task is worth 3 points, and the total is worth 7 points

	Learning task	Transfer task	Total
Pre-test	0.64 (0.88)	0.03 (0.16)	0.66 (0.95)
Post-test	2.53 (0.79)	0.89 (0.48)	3.42 (1.07)
Delayed test	1.45 (1.12)	0.35 (0.48)	1.80 (1.36)

transfer tasks, and their sum. First, the total results show that most students improved their scores from 0.66 on the pre-test to 3.42 on the post-test. No scores decreased from the pre-test to the post-test. A t-test revealed a significant difference between the pre-test and post-test at $p < .001$. In the delayed test, the average score was 1.80, which was higher than the pre-test, indicating that the percentage of correct answers was maintained.

In a previous study, Aikawa et al. (2020) conducted a laboratory-scale experiment comparing the proposed system to one with auxiliary problems but no automatic presentation function. As a result, there was a large difference in the effect size between the pre-test and post-test (1.70 for the experimental group and 0.82 for the control group). Furthermore, in the log analysis, many learners in the control group were unable to perform the appropriate auxiliary problem transition. From this, we found that not including auxiliary problems, or their including but could not be used appropriately, cannot adequately support the improvement of learning outcomes.

The problems given in the test are shown in Figures 5 and 6, and the graphs of the number of correct answers and the percentage of correct answers for each problem are shown in Figure 7. In this system, problems that were considered too difficult for junior high school students (Figure 5(d)) were implemented as original problems in Figure 4 and were practiced. However, more than half of the students answered the problems correctly during the class, and 30 answered them correctly on the test, while none answered them correctly in the pre-test. This suggests that students can learn to solve problems that they are initially unable to by using this system. Additionally, 30 of the 74 students correctly answered the most basic problem (Figure 5(a), Figure 4 support problem 7) in the pre-test, while all 74 correctly answered it in the post-test.

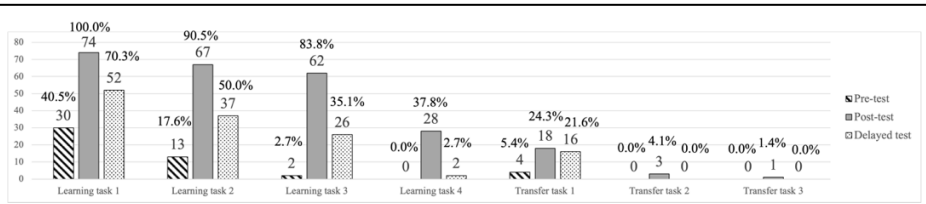


Fig. 7 Scores and the percentage of correct answers per problem in the test

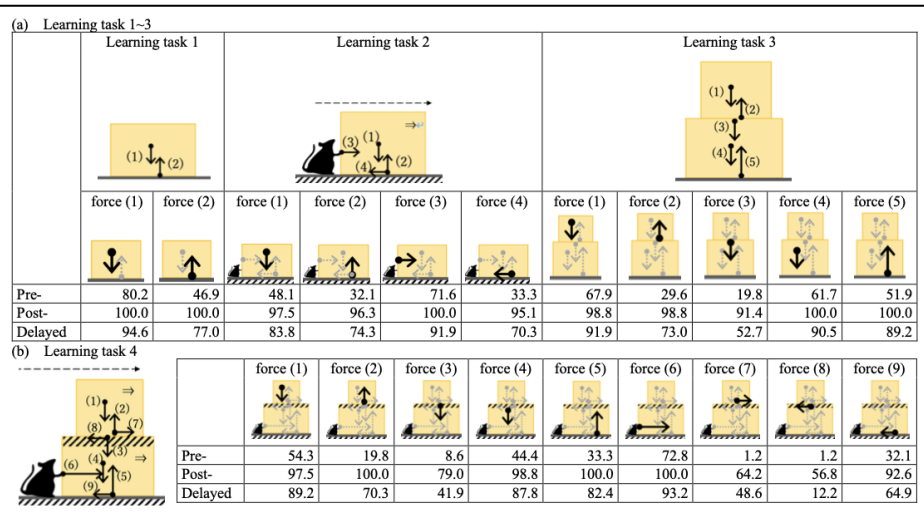
Furthermore, for Learning Task 3 (corresponding to Figure 5(c) and Figure 4 support problem 5), only two students answered it correctly in the pre-test, whereas 62 answered it correctly in the post-test. We consider it a significant achievement that the correct response rate increased from 2.7% to 83.8% between the pre-test and the post-test.

Force-by-force analysis of the test

In the "Test results" subsection, each problem was scored on a problem-by-problem basis, but here we analyze each force individually and assess whether they promote greater understanding. In the problem-by-problem scoring, the correct answer was that in which all forces were drawn exactly. For example, if forces (1) and (2) were involved in a problem, then any answer in which one or both were omitted was considered incorrect. This made it impossible to determine whether the system enabled learners to draw the correct forces in the post- and delayed tests. By clarifying whether there are cases in which forces are drawn in problems with incorrect answers, and conversely, those in which forces are not drawn, we are able to examine which auxiliary problems were effective and which need to be improved. Therefore, in the next section, we present the results of a force-by-force analysis of the test results.

Analysis of each problem

First, the forces involved in the problems are numbered in Figures 5 and 6, and their scoring results are shown in Figures 8 and 9. From Figure 8, we can see that the percentage of correct answers is high for all forces in (a) learning tasks 1-3. Specifically, the percentage of correct responses in the post-test is high for all forces and does not differ from the results in the "Test results" subsection.



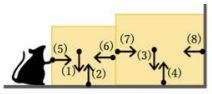
(a) Learning task 1~3

	Learning task 1		Learning task 2				Learning task 3				
	force (1)	force (2)	force (1)	force (2)	force (3)	force (4)	force (1)	force (2)	force (3)	force (4)	force (5)
Pre-	80.2	46.9	48.1	32.1	71.6	33.3	67.9	29.6	19.8	61.7	51.9
Post-	100.0	100.0	97.5	96.3	100.0	95.1	98.8	98.8	91.4	100.0	100.0
Delayed	94.6	77.0	83.8	74.3	91.9	70.3	91.9	73.0	52.7	90.5	89.2

(b) Learning task 4

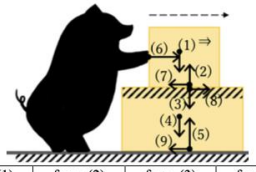
	force (1)	force (2)	force (3)	force (4)	force (5)	force (6)	force (7)	force (8)	force (9)
Pre-	54.3	19.8	8.6	44.4	33.3	72.8	1.2	1.2	32.1
Post-	97.5	100.0	79.0	98.8	100.0	100.0	64.2	56.8	92.6
Delayed	89.2	70.3	41.9	87.8	82.4	93.2	48.6	12.2	64.9

Fig. 8 Percentage of correct answers for each problem by force (learning task)



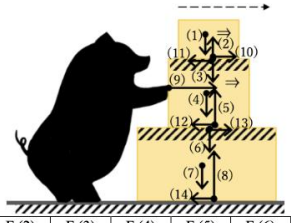
(c) Transfer task 1

	force (1)	force (2)	force (3)	force (4)	force (5)	force (6)	force (7)	force (8)
Pre-	46.9	30.9	46.9	33.3	82.7	14.8	37.0	34.6
Post-	92.6	92.6	88.9	88.9	96.3	30.9	71.6	53.1
Delayed	81.1	71.6	81.1	70.3	91.9	32.4	63.5	59.5



(d) Transfer task 2

	force (1)	force (2)	force (3)	force (4)	force (5)	force (6)	force (7)	force (8)	force (9)
Pre-	49.4	23.5	13.6	58.0	38.3	75.3	29.6	0.0	4.9
Post-	93.8	95.1	72.8	90.1	91.4	96.3	76.5	6.2	29.6
Delayed	89.2	73.0	39.2	86.5	74.3	90.5	64.9	4.1	13.5



(e) Transfer task 3
("F" is for force.)

	F (1)	F (2)	F (3)	F (4)	F (5)	F (6)	F (7)	F (8)	F (9)	F (10)	F (11)	F (12)	F (13)	F (14)
Pre-	53.1	21.0	8.6	40.7	21.0	14.8	55.6	40.7	75.3	1.2	1.2	30.9	0.0	4.9
Post-	97.5	98.8	74.1	95.1	97.5	76.5	90.1	91.4	97.5	61.7	44.4	81.5	4.9	27.2
Delayed	86.5	73.0	36.5	83.8	71.6	39.2	83.8	77.0	89.2	40.5	13.5	59.5	6.8	13.5

Fig. 9 Percentage of correct answers for each problem by force (transfer task)

Next, for learning task 4 in (b), forces (1)-(6) and (9) had high percentages of correct responses in the post-test, but forces (7) and (8) did not. Force (8) was particularly low in the delayed test at 12%, making it a primary factor in the low percentage of correct responses to that test.

Next, Figure 9 shows that for transfer task 1 in (c), forces (1), (3), and (5) had high percentages of correct responses in both the post- and delayed tests, with force (5) having a high percentage of correct responses in the pre-test. On the other hand, the response rate for force (6) was only in the 30% range for both the post- and delayed tests.

Finally, as shown in Figure 7, there were almost no correct answers for transfer tasks 2 and 3 in (d) and (e), but correct answers for forces (1), (4), and (6) in transfer task 2 and forces (1), (4), (7), and (9) in transfer task 3 exceeded 80% in both the post- and delayed tests.

Overall, the percentage of correct responses varied depending on the forces involved in the problem. Therefore, in the next section, we categorize the forces and analyze the percentage of correct responses.

Categorical analysis of forces

The forces are categorized and the percentages of correct responses are summarized in Figure 10. The categories are (a) gravity, (b) the normal force, (c-1) the force exerted by an object above, (c-2) the force exerted by an external force, (d-1) the friction force with the ground resisting a pushing force, (d-2) the friction force with an object above resisting a pushing force, (d-3) the friction force propagating to the upper object, (d-4) the frictional force propagating to the lower object, (d-5) the frictional force with the ground resisting the force propagating to the lower object, and (e) the action and reaction between objects and between objects and walls.

First, in Figure 10(a), gravity had a correct response rate of more than 80% for all problems in the post- and delayed tests. Next, (b) the normal force had a correct response rate of 80% or more in the post-test and 70% or more in the delayed test.

(a) Gravity

	Learning task 1	Learning task 2	Learning task 3		Learning task 4		Transfer task 1		Transfer task 2		Transfer task 3		
	(1)	(1)	(1)	(4)	(1)	(4)	(1)	(3)	(1)	(4)	(1)	(4)	(7)
Pre-	80.2	48.1	67.9	61.7	54.3	44.4	46.9	46.9	49.4	58.0	53.1	40.7	55.6
Post-	100.0	97.5	98.8	100.0	97.5	98.8	92.6	88.9	93.8	90.1	97.5	95.1	90.1
Delayed	94.6	83.8	91.9	90.5	89.2	87.8	81.1	81.1	89.2	86.5	86.5	83.8	83.8

(b) Normal force

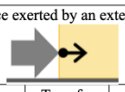
	Learning task 1	Learning task 2	Learning task 3		Learning task 4		Transfer task 1		Transfer task 2		Transfer task 3		
	(2)	(2)	(2)	(5)	(2)	(5)	(2)	(4)	(2)	(5)	(2)	(5)	(8)
Pre-	46.9	32.1	29.6	51.9	19.8	33.3	30.9	33.3	23.5	38.3	21.0	21.0	40.7
Post-	100.0	96.3	98.8	100.0	100.0	100.0	92.6	88.9	95.1	91.4	98.8	97.5	91.4
Delayed	77.0	74.3	73.0	89.2	70.3	82.4	71.6	70.3	73.0	74.3	73.0	71.6	77.0

(c) Exerted force

(c-1) Force exerted by an object above

	Learning task 3	Learning task 4	Transfer task 2	Transfer task 3	
	(3)	(3)	(3)	(3)	(6)
Pre-	19.8	8.6	13.6	8.6	14.8
Post-	91.4	79.0	72.8	74.1	76.5
Delayed	52.7	41.9	39.2	36.5	39.2

(c-2) Force exerted by an external force



Learning task 2	Learning task 4	Transfer task 1	Transfer task 2	Transfer task 3
(3)	(6)	(5)	(6)	(9)
71.6	72.8	82.7	75.3	75.3
100.0	100.0	96.3	96.3	97.5
91.9	93.2	91.9	90.5	89.2

(d) Frictional force

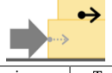
(d-1) Friction force with the ground resisting a pushing force

	Learning task 2	Learning task 4	Transfer task 2	Transfer task 3
	(4)	(9)	(7)	(12)
Pre-	33.3	32.1	29.6	30.9
Post-	95.1	92.6	76.5	81.5
Delayed	70.3	64.9	64.9	59.5

(d-2) Friction force with an object above resisting a pushing force

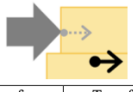
Learning task 4	Transfer task 3
(8)	(11)
1.2	1.2
56.8	44.4
12.2	13.5

(d-3) Friction force propagating to the upper object



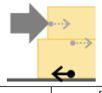
Learning task 4	Transfer task 3
(7)	(10)
1.2	1.2
64.2	61.7
48.6	40.5

(d-4) Frictional force propagating to the lower object



	Transfer task 2	Transfer task 3
	(8)	(13)
Pre-	0.0	0.0
Post-	6.2	4.9
Delayed	4.1	6.8

(d-5) Frictional force with the ground resisting the force propagating to the lower object



Transfer task 2	Transfer task 3
(9)	(14)
4.9	4.9
29.6	27.2
13.5	13.5

(e) Action and reaction

(e) Action and reaction between objects and between objects and walls

Transfer task 1		
(6)	(7)	(8)
14.8	37.0	34.6
30.9	71.6	53.1
32.4	63.5	59.5

Fig. 10 Percentage of correct answers for each category of force

Next, the percentage of correct answers for (c-1) the force exerted by an object above was less than 20% for all problems in the pre-test and more than 70% in the post-test. This indicates that the system effectively helped learners understand this force. On the other hand, the percentage of correct answers in the delayed test was lower than in the post-test. The percentage of correct answers for (c-2) the force exerted by an external force was more than 80% in both the post- and delayed tests, as high as the percentage of correct responses for gravity and the normal force.

Next, let us look at the frictional forces. The response rate for (d-1) the friction force with the ground resisting a pushing force was generally good, but those for the other frictional forces were generally low. The response rate for (d-3) the friction force propagating to the upper object was higher than that for the other frictional forces.

Discussion of test results

The correct response rates were high for (a) gravity, (b) the normal force, (c-2) an external force, and (d-1) the friction force with the ground resisting a pushing force. These forces are included in the series of links among support problems 2, 3, and 7 in Figure 4. (a) Gravity and (b) the normal force are intended to be learned in support problem 7, and (c-2) the force exerted by an external force is intended to be learned in the link between support problems 7 and 3. The link between support problems 3 and 2 is used to teach (d-1) the friction force with the ground resisting a pushing force. We believe that linking this series of problems in this way effectively helped learners understand these forces.

On the other hand, the response rates for the frictional forces other than (d-1) the friction force with the ground resisting a pushing force were generally low. Of these, (d-2) the friction force with an object above resisting a pushing force is the one intended to be learned in the link between support problems 2 and 1 in Figure 4. Therefore, we can conclude that force (8) in the original problem was not learned well because of the large difference between the support problems at both ends of this link.

The correct response rate was low for (d-2) the friction force with an object above resisting a pushing force, but that for (d-3) the friction force propagating to the upper object was comparatively high. This is the force that we aimed to have the students learn in the link between support problems 7 and 4 in Figure 4.

Furthermore, (d-4) the frictional force propagating to the lower object and (d-5) the frictional force with the ground resisting the force propagating to the lower object appeared only in the transfer problems. The percentage of correct answers to these problems was particularly low, suggesting that they were difficult for learners to understand using this problem graph.

Additionally, the correct response rate for (c-1) the force exerted by an object above increased in the post-test but decreased in the delayed test. This is the force that is intended

to be learned in the link between support problems 7 and 6 in Figure 4. However, it is not only this link but also the link up to support problem 5 that is influential. This series of links was used by the system, but there were problems in the delayed test with the assignment corresponding to support problems 7 and 5, but not to support problem 6. Therefore, we conclude that solving the series of auxiliary problems from support problems 7 to 5 was effective.

Questionnaire

The results of the questionnaire are presented below. The questions were asked using a six-point scale, with 1 corresponding to "not at all agree" and 6 corresponding to "very much agree," and the average was calculated. For the sake of clarity, the system was referred to as "application" and the auxiliary problems as "support problems" within the questionnaire. We asked (No. 1) "Do you feel that the support problems helped you solve the original problem?" and (No. 2) "Did you feel that the support problems were pertinent to your errors?"

The results showed that No. 1 received a high score of 5.46 and No. 2 received a high score of 5.06. These suggest that learners are positive that the auxiliary problem framework facilitates self-overcoming.

Log analysis

Analysis of self-overcoming

We evaluated the system logs from the learning activity to evaluate the students' understanding. Some of the results from same class practice session are reported in Aikawa et al. (2022), and the results for the delayed test and force-by-force analysis are reported in Aikawa et al. (2023). In this paper, we added an analysis of the impasse by analyzing the system logs. First, the completion rate of the exercises was 70.3% (52/74), meaning that 70.3% of the students were able to solve the original problem after failing to do so initially. Thus, we conclude that the 64 minutes of system practice time was sufficient for 70% of the students to complete the original problem after solving the auxiliary problems.

To better enable self-overcoming, a strategy was proposed that simplifies the unsolvable problems by partializing or specializing them. This allows a learner to find problems that can be solved on their own, before then revealing the differences between the solvable and unsolvable problems. The usefulness of this strategy was previously evaluated (Hayashi et al., 2014).

In this study, we attempted to support self-overcoming by presenting auxiliary problems with EBS. We investigated the degree to which our system promoted self-overcoming by analyzing its usage log. Specifically, we extracted a learner's errors to the original problem

from their answer history and compared them to the answers given once they were able to solve the original problem after working through the auxiliary problems. If one error in the previous answer history was eliminated in the next answer history, it was counted as 1 and repeated.

After solving several problems, the system analyzed the errors and presented additional auxiliary problems accordingly. Ultimately, a learner solves the same problem multiple times before moving on to another problem (unless it is answered correctly the first time), and we refer to the history of multiple solutions to a single problem as an "answer series" (Figure 11). For this study, only the answer series of the original problem was extracted and analyzed to calculate the number of times self-overcoming occurred.

The log of one learner who solved the original problem during the system practice time is given in Table 2 as an example of self-overcoming. This table summarizes the errors in the student's answer for each problem transition. The "problem number" is the problem that the student worked on, and this line contains the information for a single answer series. The "correct/incorrect" answer is the last correct or incorrect answer in the answer series for that problem (this is related to the next problem presented by the system), and "error" indicates which force in the answer series for that problem was missing and how many times. For example, the first row indicates that the original problem was being worked on, the answer was incorrect, and the system moved on to the next problem. The student made three mistakes for force (3), three for force (7), three for force (8), and three for force (9) in the answer series for the original problem. Note that in the actual exercise, students who were able to solve the original problem within the exercise time were asked to work on the

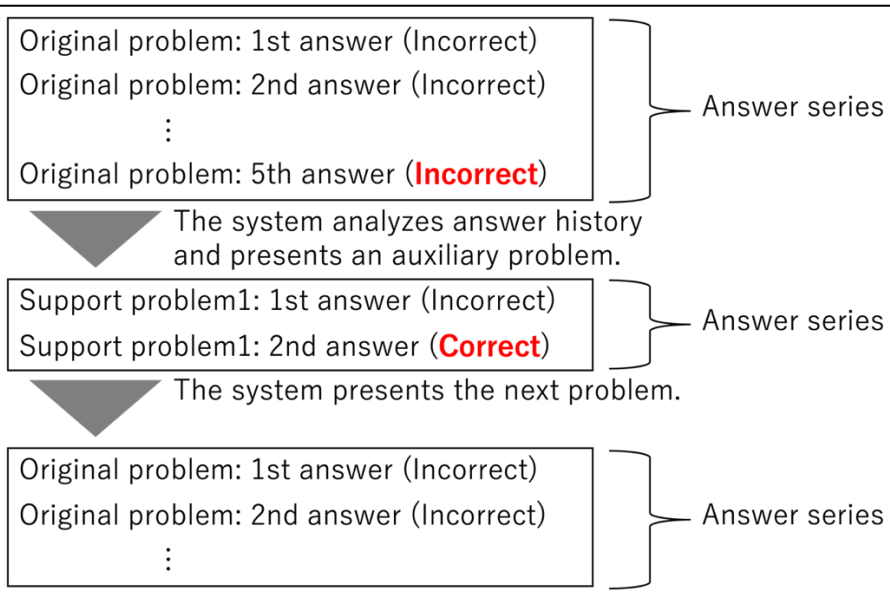


Fig. 11 An answer series

system again, but they were not included in the log analysis since they had to work on the correct answer once after the second attempt.

We can also see from Table 2 that when the students worked on the first original problem presented to them, they made three errors each for forces (3), (7), (8), and (9) as they repeated their answers. The system then presented support problems 7 and 6 in sequence for them to learn the highest priority force, force (3), since they all had the same number of exclusions in the answers. The student answered support problem 7 correctly but was unable to do so for support problem 6, so they worked on support problem 7 again, following the priority order of the errors. They then answered the problem correctly and moved on to support problems 6 and 5, in accordance with the sequence shown in the problem graph. However, when they failed to answer support problem 5 correctly, the system again referred to the error and presented support problems 7 and 6 corresponding to the most frequently missed force, force (3). The student answered support problems 7 and 6 correctly, then answered support problem 5 correctly, and again worked on the original problem. Here, the student's answers are missing the arrows for forces (7), (8), and (9), but force (3), which was missing at the beginning, no longer is. This suggests that they learned force (3) through the auxiliary problems.

However, they were still incorrect regarding forces (7), (8), and (9), so support problems 7 and 4 were presented in sequence, which allowed them to learn forces (7) and (8) in the order of highest priority. They answered each problem correctly, then worked on the original problem again according to the problem graph. In this case, the student's answers

Table 2 System logs of students who answered the original problems correctly

	Problem No.	Correct/incorrect	Error
1	Original problem	Incorrect	(8): 3 times, (3): 3 times, (9): 3 times, (7): 3 times
2	Support problem 7	Correct	(4): 1 time, (5): 2 times
3	Support problem 6	Incorrect	(3): 1 time, (4): 1 time, (5): 1 time
4	Support problem 7	Correct	
5	Support problem 6	Correct	(3): 1 time, (4): 1 time, (5): 2 times
6	Support problem 5	Incorrect	(3): 9 times, (2): 8 times, (1): 1 time
7	Support problem 7	Correct	
8	Support problem 6	Correct	
9	Support problem 5	Correct	
10	Original problem	Incorrect	(8): 13 times, (9): 8 times, (7): 13 times
11	Support problem 7	Correct	
12	Support problem 4	Correct	(7): 1 time
13	Original problem	Incorrect	(8): 10 times, (9): 1 time, (3): 1 time, (7): 1 time
14	Support problem 2	Correct	
15	Support problem 1	Incorrect	(8): 9 times, (3): 8 times, (9): 3 times
16	Support problem 2	Correct	(8): 6 times, (3): 2 times
17	Support problem 1	Correct	
18	Original problem	Correct	

were missing the arrows for forces (3), (7), (8), and (9), but since force (8) was wrong 10 times while the other forces were only wrong once each, we can infer that they did not understand force (8). Consequently, the system presented support problems 2 and 1 in that order. They answered support problem 2 correctly and made a mistake in support problem 1. They were then presented with these problems again, and this time they answered both correctly. The student then completed the exercise by correctly answering the last original problem presented.

When the system logs of 74 students were analyzed, self-overcoming was noted to have occurred 533 times for errors that were made more than once. It is possible that an error that occurred only once could have been due to a careless mistake. Therefore, for the sake of convenience, we counted the number of self-overcoming attempts for errors that were made three or more times. The number of instances of self-overcoming, then, was 424.

The number of self-overcoming errors was calculated by categorizing the errors based on force (Table 3). The number of errors was counted in the same way, and we also calculated the percentage of self-overcoming errors. Force (3) was found to have the highest self-overcoming frequency having occurred 80 times, followed by the frictional forces (7), (8), and (9) with a frequency of 60% (~70 times). Forces (1), (2), (4), (5), and (6) had relatively few self-overcoming instances, but they all had a high percentage of self-overcoming errors (over 40%), while forces (3), (7), (8), and (9) had relatively low percentages.

The students were divided into two groups, those who completed the exercises and those who did not. The results show that the mean number of self-overcoming attempts by students who completed the exercises was 5.88 with a standard deviation of 2.14, while for students who did not complete the exercises it was 5.36 with a standard deviation of 2.16, indicating that there was little difference. Additionally, when the mean of the number of self-overcoming instances for each force was calculated (Figure 12), it was found that those who completed the exercise exhibited self-overcoming less frequently for forces (1), (2), (4), and (5) and more frequently for forces (3), (7), (8), and (9). Conversely, the participants who did not complete the exercises were more likely to exhibit self-overcoming for forces (1), (2), (4), and (5), and less likely for forces (3), (7), (8), and (9).

Table 3 Percentage of self-overcoming instances and error times for each force

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of self-overcoming instances	29	38	80	24	28	23	71	69	62
Number of errors	63	96	209	55	63	49	283	302	174
Percentage of self-overcoming (%)	46	40	38	44	44	47	25	23	36

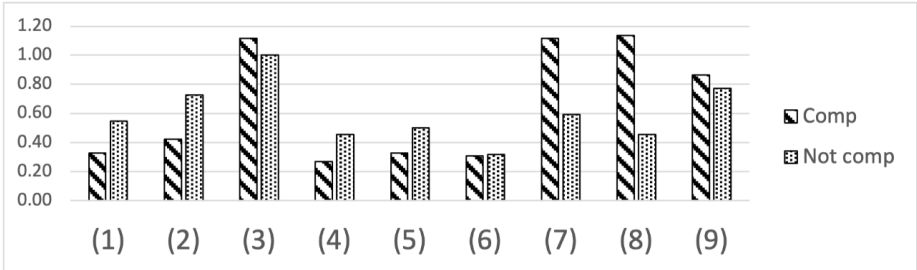


Fig. 12 Comparison of the average number of self-overcoming attempts per ability between those who completed the exercises and those who did not

Discussion of log analysis results

The total number of self-overcoming instances by the 74 students during the use of the system was 424 (based on errors made more than three times), giving an average of 5.7 self-overcoming attempts per student.

We will focus on the two groups, those who completed the exercises and those who did not, separately for the rest of this discussion. The average number of self-overcoming occurrences did not change between them.

However, from the "Analysis of self-overcoming" subsection, we found that the forces for which learners exhibited self-overcoming tended to differ between the groups. Those who completed the exercises tended to exhibit self-overcoming for forces (1), (2), (4), and (5) less frequently, but more frequently for forces (3), (7), (8), and (9). The opposite was true for the participants who did not complete the exercises.

Forces (1), (2), (4), and (5) correspond to gravity and the normal forces. Force (3) is the force exerted by the lower object on the upper object when they are stacked on top of each other, and forces (7), (8) and (9) correspond to frictional forces. From the "Categorical analysis of forces" subsection, the percentage of correct answers for forces (1), (2), (4), and (5) was high, while that for forces (3), (7), (8), and (9) was low by comparison, suggesting that the difficulty level for the students was similar. The fact that the number of errors for forces (1), (2), (4), and (5) was less than 40 while that for forces (3), (7), (8), and (9) was more than 60 supports the idea that those forces were more difficult for learners to understand.

Discussion

Comparison of results of force-by-force analysis against tests and self-overcoming analysis against logs

We now summarize and discuss the results of the test analysis and the log analysis. We focus on common problems that were handled in the test and in the system, especially learning task 4 in the test (the original problem in the system).

We compare the results of the force-by-force analysis for the test to the self-overcoming analysis for the log. First, we calculated the correlation between the percentage of correct answers per task in the pre-test and the number of errors per task in the log analysis for learning task 4. The results showed a strong negative correlation with a correlation coefficient of $r = -.85$. This indicates that a lower score in the pre-test corresponded to more errors being made in the system, which reflects the difficulty level of the associated force.

Next, we analyzed the test results to look for a correlation with self-overcoming. We differentiate between the self-overcoming count and the self-overcoming percentage. The self-overcoming count is the number of times that self-overcoming occurred, calculated in the manner specified in the "Analysis of self-overcoming" subsection. A high number for the self-overcoming count indicates that the learner was able to overcome many errors, thus leading to an increase in test scores. However, there is a caveat: since the number of errors was not considered, the same value was obtained whether the number of errors was high (high difficulty) or low (low difficulty) if the number of self-overcoming errors was the same. Consequently, we hypothesized that there was likely a correlation between the number of self-overcoming occurrences and test growth (which would be reflected in the difference in scores between the pre- and post-tests).

The self-overcoming rate is the number of self-overcoming occurrences divided by the number of errors. A high self-overcoming rate indicates that the learner was able to overcome most of their errors, allowing them to achieve a high post-test score. However, the values for the self-overcoming rate are the same regardless of whether the number of self-overcoming errors and the total number of errors are both high or low. Consequently, we suspected that there was a correlation between the self-overcoming rate and the correct response rate in the post-test. We looked for a few correlations: one between the difference in the pre- and post-tests and the number of instances of self-overcoming, one between the post-test and the percentage of self-overcoming occurrences, and one between the delayed test and the percentage of self-overcoming occurrences.

The correlation between the difference in the pre- and post-tests and the number of instances of self-overcoming was positive, $r = .43$. This indicates that a greater number of

self-overcoming occurrences in the system log corresponded to a greater increase in the test scores. The correlation between the post-test and the self-overcoming rate was strongly positive, r = .94. This indicates that a higher self-overcoming ratio in the system logs corresponded to a higher score in the post-test. The correlation between the delayed test and the self-overcoming rate was strongly positive at r = .90. This indicates that a higher self-overcoming ratio in the system logs corresponded to a higher score on the delayed test. This suggests that the results for instances of self-overcoming are reflected in the percentage of correct responses in the test.

We will now look at each force in detail for learning task 4. Forces (1), (2), (4), and (5) correspond to gravity and the normal forces, and they have a self-overcoming ratio of more than 40%. Force (6) had a similar tendency. The self-overcoming rate was high, but the number of instances of self-overcoming was less than 40. This is because the number of errors was low to begin with and the difficulty level was low. The three forces correspond to (a) gravity, (b) the normal force, and (c-2) the force exerted by an external force, respectively (see the "Categorical analysis of forces" subsection for further details), and all of them were answered correctly in the delayed, post-, and pre-tests.

Force (3) corresponds to (c-1) the force exerted by an object above. It had a self-overcoming rate of 38%, but it also had the highest number of associated instances of self-overcoming (80 times). The number of errors was also high (209 times), suggesting that it was more difficult for learners to understand than (a) gravity, (b) the normal force, and (c-1) the force exerted by an object above. These results are not very different from those for (c-1) the force exerted by an object above, for which the correct response rate was low for the pre-test and high for the post- and delayed tests.

Force (9), corresponding to (d-1) the friction force with the ground resisting a pushing force, also showed a similar trend.

For forces (7) and (8), the percentages of self-overcoming were low at 25% and 23%, respectively. On the other hand, the number of instances of self-overcoming for these forces were 71 and 69, which were higher than those for (a) gravity, (b) the normal force, and (c-1) the force exerted by an object above. This suggests that the number of errors was high, as was the difficulty level. Forces (7) and (8) correspond to (d-3) the friction force propagating to the upper object and (d-2) the friction force with an object above resisting a pushing force. These results were consistent with a lower percentage of self-overcoming and lower results for the post- and delayed tests than the other forces, and with a higher number of self-overcoming occurrences and a higher improvement in scores between the pre- and post-tests.

This suggests that the results for the instances of self-overcoming during the use of the system are reflected in the percentage of correct answers on the test.

Analysis of the pre-test divided into upper and lower groups

Since the mean score of the pre-test was 0.67, we divided the learners into two groups: an "upper group" (above average, 1 or more points on the pre-test) of 31 learners and a "lower group" (below average, 0 points on the pre-test) of 43 learners. The results are summarized in Table 4. The results of the t-test show that there is a significant difference between the two groups for the three tests. The results in Table 3 also varied between the groups, as shown in Table 5. The higher groups had higher percentages of self-overcoming for each force, but the lower groups also had high percentages, with (1), (4), (5), and (6) at 40% or higher. Those with relatively low percentages, (7) and (8), were almost identical to those of the upper group. This indicates that the system encourages self-overcoming regardless of individual differences in ability, and that it has the potential to encourage self-overcoming at an even higher rate for high-performing learners, which is also reflected in the test results.

Implications

First, our results suggest that self-overcoming occurs when using this system. Additionally, 70% of the students were able to correct the original problem after using it, which is a promising result.

Table 4 Test results divided into upper and lower groups. Values outside and inside the parentheses are mean values and standard deviations, respectively

	n	Pre-test	Post-test	Delayed test
Upper group	31	1.58 (0.83)	3.90 (0.89)	2.29 (1.14)
Lower group	43	0.00 (0.00)	3.07 (1.04)	1.44 (1.39)

Table 5 Percentage of self-overcoming instances and error times for each force divided into upper and lower groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Upper group									
Number of self-overcoming instances	7	11	35	6	6	9	31	30	21
Number of errors	14	26	81	12	10	14	121	129	55
Percentage of self-overcoming (%)	50	42	43	50	60	64	26	23	38
Lower group									
Number of self-overcoming instances	22	27	45	18	22	14	40	39	41
Number of errors	49	70	128	43	53	35	162	173	119
Percentage of self-overcoming (%)	45	39	35	42	42	40	25	23	34

One of the novelties of this study is the force-by-force analysis. This suggests that learners learn at the conceptual level of “gravity” and “vertical drag” through the framework of EBS and auxiliary problems. Prior research has not gone so far as to discuss the acquisition of this concept of force.

This also made it possible to discuss each force in the problem. In this study, the system developer designed the auxiliary problems, but we believe that the design of problem graphs is a major issue for practical use. Therefore, as directions for future research, we believe that it is necessary to develop a function by which the system can automatically generate auxiliary problems, and we believe that the above data for the number of errors and the number of self-overcoming instances for each force will enable us to create more effective auxiliary problems.

In addition, we believe the framework in this study can be applied to other subjects and instructional activities. First, EBS is not limited to mechanics, and has been developed in other fields (Kurokawa et al., 2018). EBS keeps rules for generating simulations inside the system and adds to or modifies them based on the learner’s incorrect answers, thus generating strange simulations. These are the constraints in the target problem and the rules of the domains: in the case of physics, they describe the forces acting on each object and the relationship between forces and motion. This method would cover various well-defined fields in which such rules can be explicitly described. Therefore, it is thought that the rules for the domain will be described similarly in other fields covered by EBS, and therefore, auxiliary problems can be created by decomposing those rules. Furthermore, self-overcoming analysis can also be performed by analyzing the relationship between errors and overcoming the components of rules in the target domain.

We also believe that the generation of auxiliary problems and the analysis of self-overcoming can be adapted to systems with domain models other than EBS. However, since we consider self-overcoming to be a result of auxiliary problems and trial-and-error, whether it is possible depends on whether the target learning support environment can provide trial-and-error.

Conclusions

EBS visualizes errors in learners’ answers and is effective at helping them correct those errors. There are instances, however, in which learners will get stuck during the learning process. In this study, we developed an EBS system that presents auxiliary problems to learners who get stuck during exercises, problems that are specifically adapted to the areas where they are having the most difficulties. We then conducted an evaluation experiment in a junior high school to assess the system’s effectiveness, demonstrating that it is effective for helping that demographic. The results from analyzing the test answers and system logs indicated that a lot of self-overcoming occurred during the use of the system.

In future work, we would like to extend the problem graph by implementing other problems, and to assess the effectiveness of those problems through further experiments. One shortcoming of using the system log to assess self-overcoming is that it is difficult to distinguish between intentional actions and errors due to guessing and simple mistakes from the log data. However, these kinds of errors can be eliminated to some extent by counting the ones that were made more than three times as errors retained by the learner. There are cases, though, in which a learner repeatedly answers a problem while forgetting to input the correct answer, and we are not able to accurately distinguish between this and intentional actions. We would like to address this in the future. As for accidental correct answers, the probability of describing the correct answer from the available forces in the problem is less than 0.01%, so it is highly unlikely that a learner will reach the correct answer if they do not understand the problem. It is also necessary to examine whether the 30% of learners who could not solve the original problem in the 64-minute system practice time failed to do so because they did not have enough time or because they could not have even if they did. Additionally, the current problem graph was manually generated by focusing on each force individually, but we plan to adopt a method for automatically generating auxiliary problems from original problems (Aikawa et al., 2021).

Abbreviations

EBS: Error-based Simulation; Mechanics EBS: Error-based Simulation for mechanics; ICM: Increasingly complex microworlds.

Acknowledgements

Not applicable

Authors' contributions

This research was designed by Nonoka Aikawa, Kento Koike, and Takahito Tomoto based on the discussions with Tomoya Horiguchi and Tsukasa Hirashima. The system was primarily implemented by Shintaro Maeda. The classroom practice was designed by Nonoka Aikawa, Kento Koike, and Takahito Tomoto based on the discussions with Tomoya Horiguchi and Tsukasa Hirashima and conducted with the cooperation of Isao Imai's junior high school. In addition, Tomohiro Mogi also collaborated. The analyses were primarily conducted by Nonoka Aikawa and verified by all the co-authors. Nonoka Aikawa wrote the first manuscript draft but was jointly reviewed multiple times and complemented by all the co-authors. All the authors read and approved the final manuscript.

Authors' information

AN is a Ph.D student of Tokyo Polytechnic University, Japan. Her research interests include error-based simulation.
MS is a Ph.D student of Chiba Institute of Technology, Japan. His research interests include programming learning.
MT is a Ph.D student of Chiba Institute of Technology, Japan. His research interests include programming learning.
KK, Ph.D, is a program-specific researcher of Kyoto University, Japan. His research interests include intelligent learning support systems.
TT, Ph.D, is a full professor of Chiba Institute of Technology, Japan. His research interests include intelligent learning support systems.
II is the principal of Satsukigaoka Junior High School, Japan. His research interests include Education for Science, Technology and Society.
HoT, Ph.D, is a full professor of Kobe University, Japan. His research interests include intelligent learning support systems.
HiT, Ph.D, is a full professor of Hiroshima University, Japan. His research interests include knowledge-based modeling of learning activities and design & development of learning support environments.

Funding

This work was supported by JSPS KAKENHI Grant Numbers JP22K12322, JP21H03565, and JP20H01730.

Availability of data and materials
Not applicable

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Graduate School of Engineering, Tokyo Polytechnic University, Kanagawa, Japan
[2] Graduate School of Information and Computer Science, Chiba Institute of Technology, Chiba, Japan
[3] Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan
[4] Faculty of Information and Computer Science, Chiba Institute of Technology, Chiba, Japan
[5] Chiba Municipal Satsukigaoka Junior High School, Chiba, Japan
[6] Graduate School of Maritime Sciences, Kobe University, Hyogo, Japan
[7] Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, Japan

Received: 17 April 2023 Accepted: 17 October 2023
Published online: 1 January 2024 (Online First: 13 November 2023)

References

Aikawa, N., Koike, K., & Tomoto, T. (2020). Analysis of learning activities with automated auxiliary problem presentation for breaking learner impasses in physics error-based simulations. In *Workshop Proceedings of the 28th International Conference on Computers in Education* (pp. 72–83). Asia-Pacific Society for Computers in Education.

Aikawa, N., Koike, K., Tomoto, T., Horiguchi, T., & Hirashima, T. (2021). Characterization of auxiliary problems for automated generation in error-based simulation. In S. Yamamoto & H. Mori (Eds.), *Human Interface and the Management of Information. Information-Rich and Intelligent Environments. HCII 2021. Lecture Notes in Computer Science, vol 12766* (pp. 3–13). Springer, Cham. https://doi.org/10.1007/978-3-030-78361-7_1

Aikawa, N., Maeda, S., Mogi, T., Koike, K., Tomoto, T., Imai, I., Horiguchi, T., & Hirashima, T. (2022). practical use of an error-based problem presentation system in mechanics. In *Proceedings of the 30th International Conference on Computers in Education* (pp. 118–123). Asia-Pacific Society for Computers in Education.

Aikawa, N., Maeda, S., Mogi, T., Koike, K., Tomoto, T., Imai, I., Horiguchi, T., & Hirashima, T. (2023). Analysis of Classroom Test Results for an Error-Based Problem Presentation System for Mechanics. In In H. Mori & Y. Asahi (Eds.), *Human Interface and the Management of Information. HCII 2023. Lecture Notes in Computer Science, vol 14016* (pp. 213–224). Springer, Cham. https://doi.org/10.1007/978-3-031-35129-7_15

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science, vol 7926* (pp. 431–440). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_44

Beek, W., & Bredeweg, B. (2012). Context-dependent help for novices acquiring conceptual systems knowledge in DynaLearn. In S. A. Cerri, W. J. Clancey, G. Papadourakis & K. Panourgia (Eds.), *Intelligent Tutoring Systems. ITS 2012. Lecture Notes in Computer Science, vol 7315* (pp. 292–297). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30950-2_38

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, *87*(2), 309–344. https://doi.org/10.3102/0034654316670999

Burton, R. R., Brown, J. S., & Fischer, G. (1984). Skiing as a model of instruction. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 139–150). Harvard University Press.

Hayashi, N., Shinohara, T., Yamamoto, S., Hayashi, Y., Horiguchi, T., & Hirashima, T. (2014). Scaffolding for self-overcoming of impasse by using problem simplification. In *Proceedings of the 22nd International Conference on Computers in Education* (pp. 50–58). Asia-Pacific Society for Computers in Education.

Hirashima, T., Horiguchi, T., Kashihara, A., & Toyoda, J. (1998). Error-Based Simulation for Error-Visualization and its management. *International Journal of Artificial Intelligence in Education*, *9*(1-2), 17–31.

Horiguchi, T., Imai, I., Toumoto, T., & Hirashima, T. (2014). Error-based simulation for error-awareness in learning mechanics: An evaluation. *Educational Technology & Society*, *17*(3), 1–13.

Imai, I., Tomoto, T., Horiguchi, T., & Hirashima, T. (2008). Tyugaku-rika ni okeru Error-based Simulation wo motita jugyo-zissen: “Newton ni tyosen” project [A Classroom Practice of Error-based Simulation to Improve Pupils’ Understanding of Mechanics: The “Challenge to Newton!” Project]. *Transactions of Japanese Society for Information and Systems in Education*, *25*(2), 194–203. https://doi.org/10.14926/jsise.25.194

Kurokawa, K., Tomoto, T., Horiguchi, T., & Hirashima, T. (2018). Development of a mathematical solution environment to understand symbolic expressions in mathematics. In S. Yamamoto & H. Mori (Eds.), *Human Interface and the*

Management of Information. Information in Applications and Services. HIMI 2018. Lecture Notes in Computer Science, vol 10905 (pp. 288–299). Springer, Cham. https://doi.org/10.1007/978-3-319-92046-7_26

Park, M. (2019). Effects of simulation-based formative assessments on students' conceptions in physics. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(7), em1722. https://doi.org/10.29333/ejmste/103586

Salden, R. J., Paas, F., & van Merriënboer, J. J. (2006). A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior*, *22*(3), 321–333. https://doi.org/10.1016/j.chb.2004.06.003

Shinohara, T., Imai, I., Tomoto, T., Horiguchi, T., Yamada, A., Yamamoto, S., Hayashi, Y., & Hirashima, T. (2015). Experimental evaluation of error-based simulation for dynamics problems in science class at junior high school. In *Proceedings of the 23rd International Conference on Computers in Education* (pp. 265–270). Asia-Pacific Society for Computers in Education.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Tulis, M., Steuer, G., & Dresel, M. (2016). Learning from errors: A model of individual processes. *Frontline Learning Research*, *4*(2), 12–26.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, *15*(3), 147–204.

VanLehn, K., Wetzel, J., Grover, S., & Van De Sande, B. (2016). Learning how to construct models of dynamic systems: An initial evaluation of the dragoon intelligent tutoring system. *IEEE Transactions on Learning Technologies*, *10*(2), 154–167. https://doi.org/10.1109/TLT.2016.2514422

Wenger, E. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann.

Wilcox, B. R., Pollock, S. J., & Bolton, D. R. (2020). Retention of conceptual learning after an interactive introductory mechanics course. *Physical Review Physics Education Research*, *16*(1), 010140. https://doi.org/10.1103/PhysRevPhysEducRes.16.010140

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100.

Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research and Practice in Technology Enhanced Learning (RPTEL)
is an open-access journal and free of publication fee.