Title
Which Leads to More Effective Learning in Intelligent Tutoring Software: Effort-based or Performance-based Feedback?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Sato, Shintaro
Platz, Melanie
Nagashima, Tomohiro

Publication Date
2024

Peer reviewed

Which Leads to More Effective Learning in Intelligent Tutoring Software, Effort-based or Performance-based Feedback?

Shintaro Sato (shsa00005@stud.uni-saarland.de)
Melanie Platz (melanie.platz@uni-saarland.de)
Tomohiro Nagashima (nagashima@cs.uni-saarland.de)
Faculty of Mathematics and Computer Science, Saarland University, Germany

Abstract

Feedback, when used successfully, supports student learning and motivation. Although various types of feedback are used in the actual classroom, however, most interactive learning systems provide feedback that addresses learner performance only (e.g., correctness feedback). We developed two versions of an Intelligent Tutoring System (ITS) for learning ratio calculations using mathematics number lines that differed in the type of feedback it provides: effort-based and performance-based feedback. We conducted a school-based experiment with 5th graders in Japan to test the effectiveness of the two types of feedback on student learning and motivation. The results indicate a trend that performance-based feedback in the ITS had a positive impact on student learning but no difference was found on learner motivation. This study adds new knowledge of what types of adaptive feedback are effective for student learning in mathematics.

Keywords: Feedback; Intelligent Tutoring System; Mathematics; Learning; Problem solving

Introduction

Feedback Strategy for Learning

Feedback is a pedagogical technique for promoting learning (Hattie, 2009). In a school classroom, teachers monitor students' learning progress and provide timely feedback to improve student learning and their motivation.

Past studies investigated the effectiveness of feedback on student learning (e.g., Wisniewski et al., 2020), especially in mathematics (Fyfe et al., 2023; Fyfe & Rittle-Johnson, 2016). In most studies that verify feedback to be effective for children, feedback is given immediately after solving a problem or task (Fyfe et al., 2023). However, most of these studies focus on feedback that rewards students' achievement or performance (e.g., whether a student's answer is correct or not). Considering that feedback can also target other aspects of student learning, including—but not limited to—their problem-solving struggles and learner motivation (Marbibi et al., 2022), it is worthwhile to investigate the effectiveness of feedback that not only focuses on learner performance or achievement but on other important aspects during learning.

A critical aspect of student learning that may lead to, and is closely connected to, achievement and performance is students' effort (i.e., how much effort a student puts into solving the problem). Regardless of the outcome of their problem-solving performance, feedback addressing student effort could help students learn and be motivated. For instance, when a student makes mistakes several times and feels overwhelmed, rewarding the learning process (despite the outcome) could motivate students to try again with different strategies and learn from the *failure*. On the other hand, if the student only receives feedback that targets whether their answer is correct or not, they may be demotivated to continue learning (Hau & Salili, 1996).

Indeed, research on growth mindset (i.e., the belief that one's abilities can be developed through hard work) suggests that such effort-based feedback approach could be effective. Students with a growth mindset tend to tackle more challenging tasks, learn more, and get better grades (Blackwell et al., 2007). In several studies, praising effort affects the growth mindset more positively than praising ability or intelligence (Kakinuma et al., 2021; Zarrinabadi et al. 2021). Also, praising effort has a more positive impact on students' motivation than praising their intelligence (Mueller & Dweck, 1998). These studies suggest that feedback that praises student effort may have a positive effect on their learning and motivation.

Feedback in Interactive Learning Systems

Interactive learning environments are one type of learning environments in which feedback research has been conducted extensively. When using digital learning systems, feedback can be adapted and personalized to individual students, their progress, and their outcomes (Deeva et al., 2021).

For instance, Intelligent Tutoring Systems (ITSs) are computer-based learning systems that offer learners adaptive instruction and immediate, targeted feedback (Dāboliņš & Grundspeņķis, 2013). Research has shown that ITSs can be used effectively to enhance student learning through their adaptive, step-by-step instructional support (Kulik et al., 2015). A study showed that personalized feedback in an ITS leads to a significant improvement in the learning outcomes of students (Kochmar et al., 2020). Implementing appropriate feedback in ITSs is important for developing an optimal computer-based teaching method. Still, most ITSs in the literature use feedback that addresses students' performance and achievements and not on other aspects of learning (e.g., Aleven & Koedinger, 2002; Nagashima et al., 2022).

The Learning Context: Number Lines in Mathematics Learning

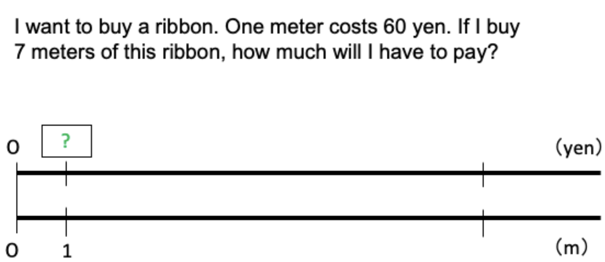
Number lines are an instructional method that visualizes relations between different quantities of a math problem

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

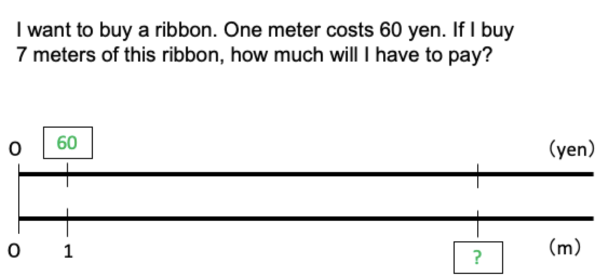
using a visual representation of numbers on a straight line. Also commonly used is the *double number line* method, which has two parallel lines representing two different units (Figure 1). The method is widely used as a representation in mathematics lessons in schools, particularly in Japan, covering a wide range of mathematics topics, including ratio calculation, basic arithmetic, and equation solving. However, it has been reported that some students do not understand the meaning of the visual representation correctly and that it is difficult for them to generate a formula/equation from the number line when working on story problems (Ishida, 2022). Another practical challenge of using number lines in classroom teaching is that drawing parallel (for double number lines), clean number lines (which many teachers and students tend to do) can take a lot of time.

Adaptive learning technologies can offer an effective and efficient way of using number lines for learning mathematics. For instance, adaptive learning technologies provide step-by-step problem-solving opportunities required for using number lines to solve word problems (see Figure 1). Also, learning technologies would save time of teachers and students for drawing number lines. A similar method using "tape/bar diagrams" has been reported effective in enhancing student learning in mathematics (Booth & Koedinger, 2012; Shirai, 2017), especially when integrated into an interactive learning systems with step-by-step guidance and feedback (Nagashima et al., 2022).

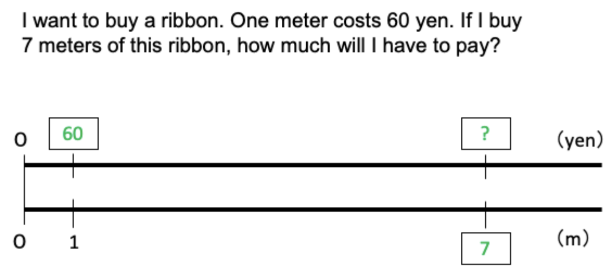
We chose ratio calculations with number lines as our learning domain because of the reported difficulty of using number lines for story problems and the step-by-step problem-solving opportunities that it provides (hence more opportunities for receiving feedback from the system).



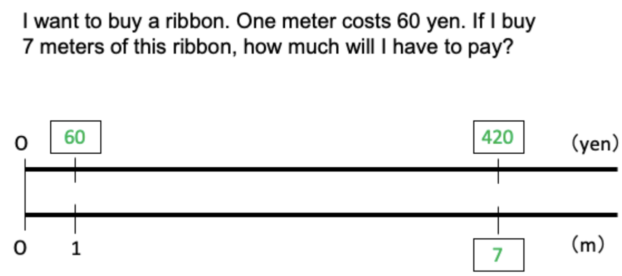
(a) Identifying the unit value.



(b) Identifying the value on the number line from the problem text.



(c) "60" multiplied by "7"



(d) The answer is "420".

Figure 1: Examples of double number lines. Students would solve these problems in a step-by-step fashion, first by identifying what the unit value is (a), then identifying how much of the unit value they are asked to multiply by (b), and then by calculating how much it would be when multiplied by the number 7 (c).

Present Study

In this study, we investigate the following research question: *Does effort-based feedback on ITS enhance learning and positive attitude in mathematics compared to performance-based feedback?* We hypothesize that students who receive effort-based feedback learn better and get more motivated to learn mathematics than those who receive performance-based feedback in the ITS for learning mathematics using number lines.

Method

Participants

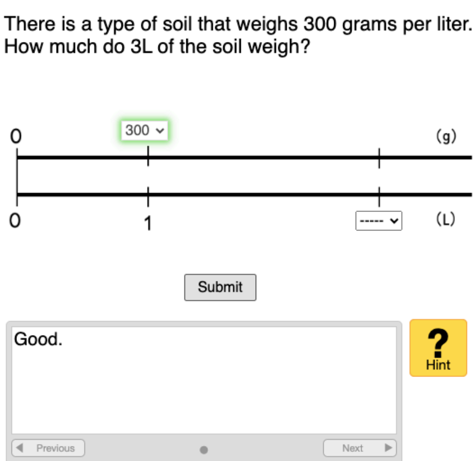
We conducted a randomized controlled experiment in a public elementary school in Japan in October 2023. Participants included 5th graders (N = 91, aged 10-11) from three different classes at the school. The students had been taught the number line method before in the class and therefore had some familiarity with the method.

Materials

Intelligent Tutoring System with Effort-based and Performance-based Feedback Messages We developed two versions of an ITS for learning ratio calculations. One gives *effort-based feedback* while the other gives *performance-based feedback*—more typical for ITS—in Japanese. For instance, as seen in Figure 2, *performance-based feedback* may say, "Nice, it is correct!" and "Not quite correct," focusing on students' problem-solving performance (i.e., correctness). On the other hand, the version of the ITS with *effort-based feedback* may say, for example, "Wow, you've been working hard!" and "Don't worry, learning is a process!", focusing more on the effort the learner has put into the task and the learning process (Figure 3) without explicitly telling whether the learner input is correct or not. In both versions, students receive feedback after pressing the "Submit" button for each step of the question. Other example feedback messages implemented in these two versions are shown in Table 1. These two versions of the ITS differed only in the type of feedback messages it gives to students. Other features, including when feedback messages are triggered (i.e., after every problem-solving step), the math problems assigned, the number of steps required to complete problems, and hint messages were consistent across the two versions.

In these versions, we implemented 11 ratio calculation problems with double number lines which differed in their difficulty (Table 2). The problem context becomes more complex and difficult for students to understand the meaning of the number line. Specifically, based on our review of the 5th-grade mathematics textbook in Japan, we created three easy problems assigned at the beginning, three medium-difficulty problems, and three difficult problems after completing the easy and medium-difficulty problems. Finally, two advanced problems that were beyond the textbook level were assigned at the end. In creating the software, we consulted with participating school's teachers regarding the appropriateness of the 11 problems and the functionality of the ITS.

Correct



Incorrect

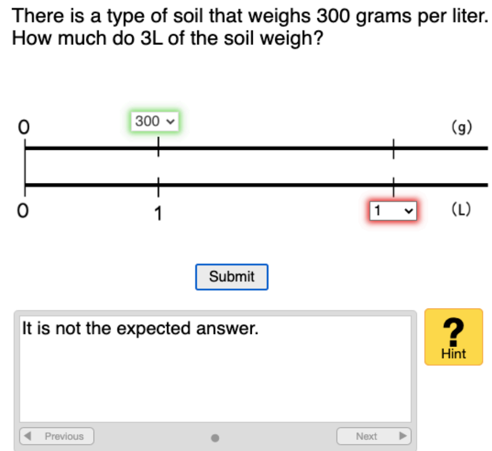
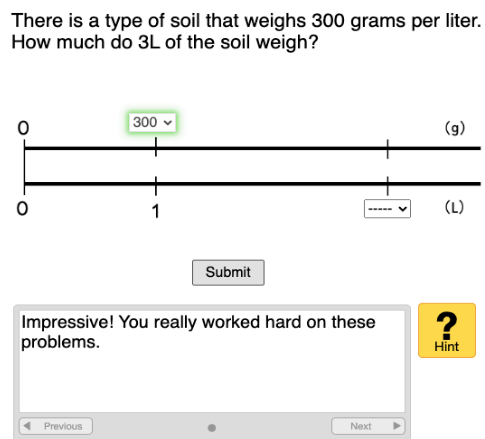


Figure 2: Examples of the ITS with *performance-based feedback* (in English).

Correct



Incorrect

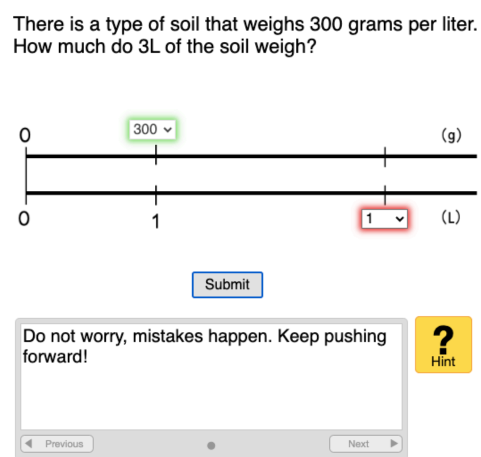


Figure 3: Examples of the ITS with *effort-based feedback* (in English).

Table 1: Example feedback messages given when a student's input is correct and incorrect.

Effort-based
【Correct】 ・Wow, you've been working hard! ・Impressive! You really worked hard on these problems. ・You've clearly made an effort to understand and solve these problems. 【Incorrect】 ・Dedication and perseverance is important for learning. ・No problem, you are still learning. ・Don't worry, learning is a process!
Performance-based
【Correct】 ・Good job. ・OK. ・Nice, it is correct! 【Incorrect】 ・Not the correct solution. ・Sorry, this answer is wrong. ・Incorrect.

Table 2: Examples of problem texts which students solved on the ITS.

Problem Text (Examples)
【Easy】 I want to buy a ribbon. One meter of a ribbon costs 60 yen. If I buy 7 meters of this ribbon, how much will I have to pay? 【Medium】 6 colored pencils cost 360 yen; how much does each pencil cost? 【Difficult】 Car A travels 200 km in 2 hours and car B travels 270 km in 3 hours. Car C travels 300 km in 6 hours. Which car is the fastest? 【Advanced】 North town potatoes are 648 yen for 9 potatoes. West town potatoes are 365 yen for 5. South town potatoes are 476 yen for 7 potatoes, and East town potatoes are 462 yen for 6 potatoes. Which town has the cheapest price per potato?

Pretest and Posttest We developed a web-based pretest and posttest to assess students' knowledge of ratio calculations with double number lines, delivered via Microsoft Forms (which the participants were familiar with, according to their teachers). The test contained 14 problems (max score: 14), and these items asked students to give specific numbers as their answer to the given problem. All problems were accompanied with double number lines. Depending on the correctness, either 1 or 0 point was given, and then summed up to make each student's score. These tests were created in consultation with the teachers of the participating school to ensure that the test items were appropriate in terms of their difficulty and familiarity with the terms used. We developed two isomorphic versions of the test that varied only with respect to the specific numbers used in the items; participants received one form as a pretest and the other as a posttest, with versions counter-balanced across subjects.

Questionnaire on student motivation towards learning mathematics To evaluate how the ITS intervention affects students' motivation towards learning mathematics, we implemented a web-based (via Microsoft Forms) questionnaire asking participants about their attitudes towards learning mathematics.

The questionnaire consisted of 15 items across four dimensions (Intrinsic Value, Self-regulation, Self-efficacy, and Test anxiety), which were selected from the Mathematics Motivation Questionnaire (MMQ) (Fiorella et al., 2021). The questionnaire used the five-point Likert scale ("never", "rarely", "sometimes", "usually", and "always").

Procedure

Students in each of the three classes (class 1: N = 31, class 2: N = 30, class 3: N = 30) were randomly assigned to either the effort-based feedback condition or the performance-based condition (Figure 4). The study was conducted in the school classroom, except for learning with the ITS, which was carried out as a homework assignment. On the first day, students worked on the pretest in their classroom for around 15 minutes. On the second day, the experimenter showed students in both conditions how to use the ITS. Students then worked on the ITS problems at home for about 7 days (the learning time was not controlled; students could work on the software as much as and as long as they wanted). Note that we decided to assign ITS problems as homework during the study (after the pretest) as there was a network problem in the school building. If students could not work on the experiment at home, they worked on it at the school with their classroom teacher.

All students had a tablet device provided by the Board of Education in the city. Students had used the device in daily classroom activities and homework.

The permission to conduct research in the school was obtained from the principal responsible for the supervision of that school and the Ethical Review Bord (ERB) of the Department of Computer Science at Saarland University (No. 23-07-3). Also, the school principal informed the parents of the participating students about the experiment and provided an opportunity to opt-out of the data collection (but none opted out).

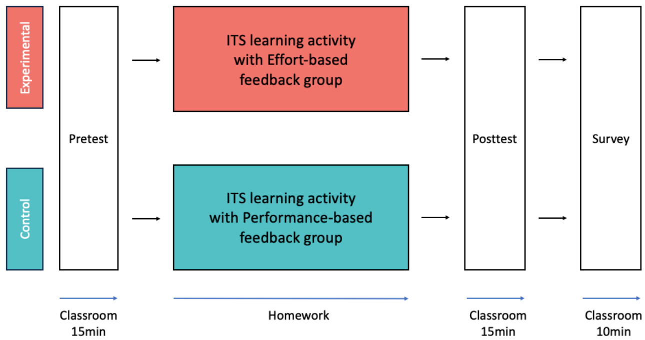


Figure 4: Study procedure

Results

In this section, we report the effect of the intervention and conditions on the measures of pretest, posttest, and Mathematics Motivation Score.

Of the 91 students who participated in the study, 65 completed all the study components. Also, of these 65 students, six scored the full score on the pretest and were excluded from the sample. Therefore, the final sample consisted of 59 students (Effort-based condition: N = 30, Performance-based condition: N = 29).

A two-sample t-test showed no significant difference on the pretest between the two groups (Effort-based: $M = 8.60$, $SD = 3.91$), Performance-based: $M = 8.17$, $SD = 3.99$).

Results on Student Learning

We compared the pretest ($M = 8.39$, $SD = 3.92$) and posttest ($M = 9.37$, $SD = 4.16$) scores of all students. First, a repeated measures ANOVA showed a significant pretest-posttest learning gain ($F(1, 58) = 4.43$, $p = .04$) for all students, regardless of the condition.

We also performed a multiple linear regression to test the effect of the intervention when controlling for the pretest and the time spent on ITS (Latency). The model had posttest as the dependent variable and the condition as the independent variables. Pretest and Latency were included as co-variates to control for the prior knowledge students had before the study and time spent on learning with the system (Note: time spent on learning with the system differed from student to student as they worked on the ITS as part of their homework activities). The model showed a trend indicating that students in the performance-based feedback condition learned more from pretest to posttest than their peers in the effort-based condition ($\beta = 1.63$, $t(52) = 1.80$, $p = .08$). Also, no significant effect of Latency was found on learning gains ($\beta = 0.0007$, $t(52) = 0.81$, $p = .42$).

Results on Students' Learning Processes

We did not find a significant difference between the conditions regarding Latency (effort-based: M = 20 minutes 09 seconds, SD = 7 minutes 16 seconds, performance-based: M = 18 minutes 28 seconds, SD = 9 minutes 21 seconds), $t(54) = 0.46$, $p = .74$. The tutor log data shows a general trend that as the level of the problem increases, the time spent on the problem also increases (Figure 5).

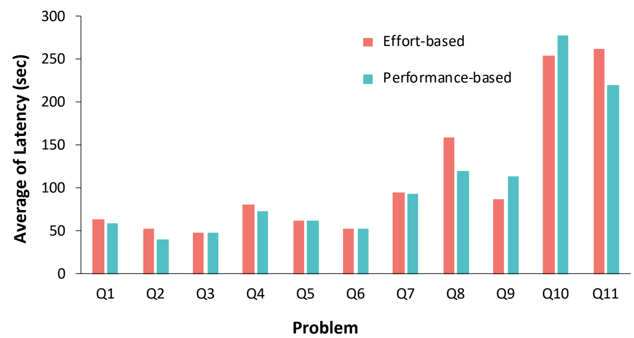


Figure 5: Students' average time spent on solving each problem in the ITS.

We also investigated whether there was any difference in the average number of incorrect attempts made in the ITS between the conditions. A two-sample t-test shows that there was no significant difference in the average number of incorrect attempts per problem in the ITS between the effort-based feedback group ($M = 0.48$, $SD = 0.43$) and the performance-based feedback group ($M = 0.60$, $SD = 0.90$), $t(57) = 0.69$, $p = .49$. Also, as a general trend, as seen in Figure 6, students rarely made mistakes; for most of the problems, students made less than one incorrect attempt (each problem had 4-9 problem-solving steps where students could have made mistakes).

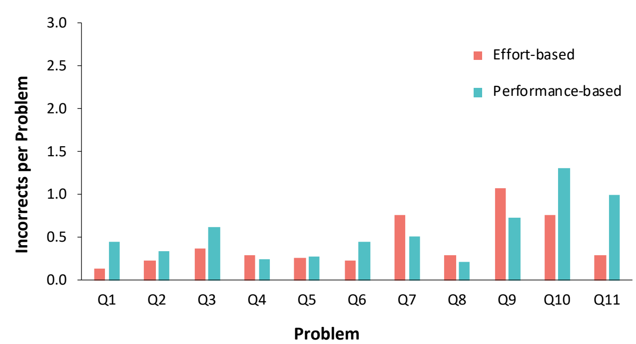


Figure 6: Average number of incorrect attempts per problem.

Attitudes in Mathematics

A two-sample t-test on the Math Motivation Score of students revealed no significant differences between the effort-based feedback group (M=3.19, SD=0.77) and the performance-based feedback group (M=3.31, SD=0.80), $t(57) = 0.61$, p = .54.

Discussion

In the current study, we investigated how the different types of feedback messages influence student learning and attitude toward mathematics using an Intelligent Tutoring System.

First, we found that students in both conditions showed a significant increase from pretest to posttest, indicating that the ITS helped students learn how to solve ratio calculations using number lines, regardless of the type of feedback students received.

Regarding the intervention effect, although it did not reach a statistical significance, the finding showed a trend indicating that the performance-based feedback group learned more. There are several speculations that we can make on why students who received effort-based feedback did not learn as much as we had predicted. First, as the log data results show that they solved problems in a relatively short time without making many mistakes, it may be that the math problems used in this experiment might have been too easy for the students (despite our consultation with teachers and math textbooks beforehand). In other words, it could be that there were only a few situations in which students in the effort-based feedback group felt challenged during the learning in ITS, and therefore their effort might not have been made enough for it to be the effective target of the feedback.

Moreover, the ITS allowed students to select answers in a multiple-choice format, which could have made the task easier than we wanted it to be. We speculate that effort-based feedback might have been more effective when students received it after making an incorrect attempt, than when they answered the step correctly (e.g., both types of feedback may have sounded positive for correct attempts, but only the effort-based feedback may have sounded positive for incorrect attempts, see Table 1). Therefore, preparing harder tasks might help students gain more from the ITS with effort-based feedback.

Further, it could be that the messages in the effort-based feedback are more ambiguous for students than those in the performance feedback. For instance, "You are still in the process of learning," an example of effort-based feedback for incorrect inputs, could also be used for the correct version (although we did not do so) because it does not tell whether the answer is correct or not. This ambiguity may have confused the students, making it hard to understand whether their input was correct or not. On the other hand, students may have found performance-based feedback more straightforward and easier to understand.

We also investigated the effect of feedback type on motivation to learn mathematics. The result shows no significant difference between the conditions. Looking at the literature, a meta-analysis found that feedback is more effective for cognitive and physical outcomes than motivational/behavioral measures (Wisniewski et al., 2020). On the other hand, such a finding often comes from a study with an intervention that lasts longer than ours (e.g., with 188 students over a two-week period of intervention where motivational messages in ITS significantly reduced students' frustration, Rajendran et al., 2019). Therefore, to reveal the relationship between motivation and feedback, it may require long-term interventions, not just a one-time brief study time, as in this study.

We recognize some limitations of the study. First, it should be mentioned that students worked on the ITS activity as their school homework at home. This means that there may have been a variety of external factors that could have influenced the results (e.g., help of parents). Future studies could administer the entire experiment at a school classroom. Furthermore, the ITS activity/intervention was only for approximately 20 minutes in this study. Thus, future research should be conducted for a long term with multiple intervention sessions using the ITS in a more controlled setting.

This study contributes empirical evidence of the effectiveness of Intelligent Tutoring Systems (ITS) and feedback strategies in actual school settings. Specifically, we found that an ITS with number lines is effective for student learning, regardless of its feedback types. Our newly-developed ITS with effort-based and performance-based feedback can be applied to other learning domains. Future studies with different content, task difficulty, school grade, and group can be conducted. If the ultimate goal of education is to support learners to become independent, then it is crucial for them to acquire the attitude to keep trying even when they face challenges. Intervention of praising effort focuses on the process of behavior, and it can allow learners to experience the value of tackling challenges in a positive manner. Therefore, it is essential that we continue to research and identify the factors and conditions that make an effort-based message effective in achieving the goal.

Acknowledgments

We would like to express our gratitude to the Japanese students who participated in our research, as well as their parents and teachers.

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science*, 26(2), 147-179.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246-263.

Booth, J. L., & Koedinger, K. R. (2012). Are diagrams always helpful tools? Developmental and individual differences in the effect of presentation format on student problem solving. *British Journal of Educational Psychology*, 82(3), 492-511.

Dāboliņš, Jānis & Grundspeņķis, Jānis. (2013). The Role of Feedback in Intelligent Tutoring System. *Applied Computer Systems*. 14. 88-93.

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & Weerdt, J. D. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162.

Fiorella, L., Yoon, S.Y., Atit, K. *et al.* Validation of the Mathematics Motivation Questionnaire (MMQ) for secondary school students. (2021). *International Journal of STEM Education*, 8, 52.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*(1), 82–97.

Fyfe, E. R., Borriello, G. A., & Merrick, M. (2023). A developmental perspective on feedback: How corrective feedback influences children's literacy, mathematics, and problem solving. *Educational Psychologist*, *58*(3), 130–145.

Hattie, J. (2009). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement (1st ed.). Routledge.

Hau, K. T., & Salili, F. (1996). Motivational effects of teachers' ability versus effort feedback on Chinese students' learning. *Social Psychology of Education*, 1(1), 69-85.

Ishida, J. (2022). A Study on Drawing and Utilization of the Number Line Diagram and the Relation Diagram in Solving Decimal Multiplication and Division Word Problems. *Journal of Science Education in Japan*, 46(4), 456–468.

Kakinuma, K., Nishiguti, F., Sonoda, K., Tajiri, H., & Tanaka, A. (2021). Effort-focused praise between friends: Effects on mindset and motivation of giver and receiver. *Social Behavior and Personality*, 49, 1-14.

Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2020). Automated Personalized Feedback Improves Learning Gains in An Intelligent Tutoring System. *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II*, 12164, 140–146.

Kulik, James & Fletcher, J. D. (2015). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*.

Marbibi, D., Añover, H., Gerilla, N., Lapidario, T., T., Mabag, M., Gerez, E., Lagahit, A., & Torreros, M. (2022). Teachers' Feedback: Influence on First Year Social Studies Students' Motivation towards New Normal Learning in Leyte Normal University. *International Journal of Social Science and Human Research*, 5(7), 2995–3002.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology*, *75*(1), 33–52.

Nagashima, T., Ling, E., Zheng, B., Bartel, A. N., Silla, E. M., Vest, N. A., Alibali, M. W., & Aleven, V. (2022). How does sustaining and interleaving visual scaffolding help learners? A classroom study with an Intelligent Tutoring System. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society (CogSci2022).* Cognitive Science Society.

Rajendran, R., Iyer, Sridhar, & Murthy, S. (2019). Personalized Affective Feedback to Address Students' Frustration in ITS. *IEEE Transactions on Learning Technologies*, *12*(1), 87–97.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10.

Zarrinabadi, N., Lou, N., & Darvishnezhad, Z. (2021). To praise or not to praise? Examining the effects of ability vs. effort praise on speaking anxiety and willingness to communicate in EFL classrooms. *Innovation in Language Learning and Teaching*, 17, 88 - 101.