International Journal of Engineering Education Vol. 39, No. 3, pp. 653–667, 2023
Printed in Great Britain

0949-149X/91 $3.00+0.00

Impact of Prompting Engineering Undergraduates to Reflect on Their Problem-Solving Skills*

SHIMA SALEHI, KAREN D. WANG, MICHAEL FLYNN and CARL WIEMAN
Stanford University, Graduate School of Education, 485 Lasuen Mall, Stanford, CA 94305-3096, USA. E-mail: salehi@stanford.edu

While learning effective problem-solving is an important goal of engineering education, "how should we teach problem-solving to engineering students?" is an ongoing challenge. In our previous works, we identified the main practices involved in solving a novel technical problem involving electrical circuits. Among these practices were reflective practices that regulate the problem-solving process by making more intentional and informed decisions. Expanding on that work, we examine in this study whether we can improve students' problem-solving by prompting their use of reflective practices. The study presented here consists of two experiments. The first experiment was conducted with 16 undergraduate students in a mechanical engineering course. Students were introduced to problem-solving reflective practices and then received prompts to engage in these practice as they worked on their weekly projects. The quality of their problem-solving was evaluated pre- and post-course using interactive electrical circuit problems embedded in an educational simulation. The improved performance in problem solving in the *mechanical engineering context* was observed to transfer and to improve problem-solving *in the context of the electrical circuit problems*. We conducted the second follow-up experiment to confirm that this improvement was the result of the prompted reflection, not simply repeated practice on the test and/or learning about electrical circuits in between the pre- and post-course evaluation. For the second study, 70 undergraduate students were randomly assigned to one of two conditions: practice solving electric circuit problems with prompted-reflection (PR) or receiving extra repeated-practice (RP) solving such problems, but without reflective prompts. Student problem-solving in the PR condition improved nearly twice as much as in the RP condition. Overall, the results of the study show that prompting students to reflect on their problem-solving produces problem-solving benefits greater than repeated practice and these benefits transfer across disciplines.

Keywords: problem-solving; reflective practices; reflective prompts; repeated practices

1. Introduction

Effective engineering education goes beyond teaching content knowledge and includes training problem-solvers who can use their knowledge to solve novel problems. Over the past two decades, the engineering and science education communities [1–8] have acknowledged the significance of training good problem-solvers and called for including teaching problem-solving practices in the curriculum. However, there remain the essential questions: "What are the characteristics of good problem-solving?" and "How can problem-solving be taught?" [9, 10].

In the classic literature on problem-solving, a problem has been defined as a goal-oriented task for which a set of required actions to reach the desired goal is not known in advance [11–14]. Extending this definition, we characterize problems in engineering and science domains as goal-directed tasks that: (1) require employing scientific and engineering knowledge; (2) the set of actions needed to reach the goal is not known in advance; and (3) encompass multiple parts and has multiple possible solution paths.

In our previous works [15], we examined problem-solving processes of individuals with a wide range of backgrounds as they worked on novel problems, which they have not seen before, in an interactive simulation to figure out a hidden combination of electrical components. We qualitatively analyzed the problem-solving of these individuals to identify the main problem-solving practices they employed, and we identified a set of practices that distinguish effective and ineffective problem-solving and characterized different levels of strength in each of these practices. This work resulted in an empirical framework for characterizing, assessing, and teaching problem-solving. The problem-solving framework has eight main practices divided into two categories: execution practices and reflective practices. Execution practices are what a problem-solver does to solve a problem and include: problem definition and decomposition of the problem into suitable sub-problems, data collection, data recording, data interpretation. Reflective practices are how a problem-solver decides what to do to solve a problem and include: reflection on problem definition and assumptions, reflection on knowledge (what is known and what needs to be known), reflection on effectiveness of strategy being used to solve the problem, reflection on solution, including testing and verification of the solution. In other words, reflective practices are regulative processes

* Accepted 8 February 2023.

in which a problem-solver engages to make intentional and informed decisions about what to do to solve a problem. These four reflective practices were further validated by Price et al. [16], when they interviewed a series of scientists and researchers to identify decisions they engage in while solving a problem.

While, many but not all problem-solving practices of the above problem-solving framework have been noted and discussed in previous works related to engineering education, for example: problem definition and decomposition of the problem into suitable sub-problems [17]; data collection [18]; reflection on knowledge (what is known and what needs to be known) [19]; and reflection on solution, including testing and verification of the solution [20]; the reflective practices and the teaching of them have received relatively little attention. As Turns et al. [21] stated, "Although reflection has not historically received great attention in engineering education, the recent efforts call for more emphasis on the role of reflection in engineering education. Ambrose [22] calls for curriculum change by arguing that "students learn by doing, but only when they have time to reflect – the two go hand in hand. Why, then, don't engineering curricula provide constant structured opportunities and time to ensure that continual reflection takes place?" (p. 1).

The current work presents two experiments that examine whether it is possible to teach students the reflective practices identified in [15] by providing them with reflective prompts as they work on novel problems. For example, students were prompted to reflect on their problem-solving strategies by being asked to consider "different strategies for solving the problem", and to consider "criteria to compare and contrast these different strategies." Previous studies have shown the benefits of reflective training in contexts such as reading [23], programming [24–26] mathematical reasoning [27, 28], and engineering [29, 30]. In other cases, researchers considered the questions posed during peer feedback sessions as prompts to encourage reflection [31]. The reflective trainings in such studies were quite general. Students were encouraged to think about questions such as: what were they doing? why were they doing it? and whether they should change what they were doing. The effectiveness of such general reflective training highly depends on the quality of questions posed by students and/or their peers. In the study presented here, we provided the students with a set of specific, and clear reflective prompts, asking them to think about how they could improve on specific problem-solving practices, as listed above. We then measured if their problem-solving did improve as a result of these prompts as they engaged in solving problems. Providing students with specific and clear reflective prompts not only encourages student reflection during engineering problem-solving, but also makes sure the effectiveness of reflection does not depend on students' ability to pose helpful questions during reflection.

2. Research Questions

The research questions for this study are:

RQ1. Can students' problem-solving improve by prompting reflective practices during problem solving?

RQ2. Do the benefits of learning reflective practices in one context transfer to different contexts?

RQ3. Does reflection increase the probability of students correctly solving the problem?

RQ4. Would students perceive the reflective prompts to be helpful?

These research questions were examined through two experiments. In the first experiment, we examined whether receiving reflective prompts during mechanical engineering course projects would improve students problem-solving as measured by pre- and post-course tests. In the second study, we compared the effect of reflective prompts with repeated practice of similar problems across two conditions: in one condition students solve two electricity problems and received reflective prompts in between, while in the other conditions they practiced solving three similar problems. The details of methods and participant population is presented in the following.

3. Methods

3.1 Problems Used in the Experiments

To study problem-solving, we needed a family of similar novel problems such that they are sufficiently unstructured to call upon a variety of problem-solving practices, but also are constrained enough to be solvable in a fraction of an hour and allow problem-solving methods to be readily compared. These novel problems also needed to be challenging across different experience levels in the problem domain: not to be a routine exercise for experienced problem-solvers, nor be impossible to solve for less experienced problem-solvers. The problems also needed to be authentic and to call on content knowledge in a domain, but that knowledge should be reasonably familiar to a wide-range of problem-solvers. We designed a family of problems about simple electrical circuits, embedded in PhET Circuit Construction Kit (CCK) interactive simulation environment [32, 33] to satisfy these criteria. In these problems, a black box hides an electrical

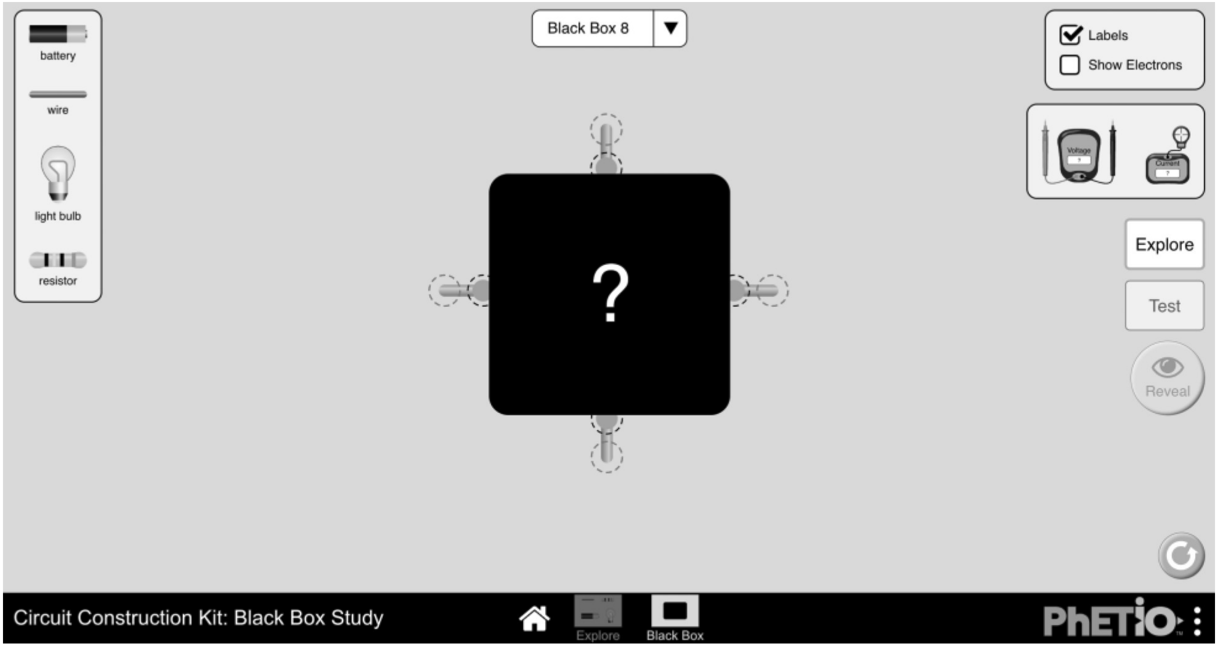

Fig. 1. Illustration of the black box problem. The back box is hiding an electrical configuration, and the task of the problem-solver is to use the tools provided in simulation interface to identify the hidden configuration. Each pair of four wires protruding from the black box is either: (1) not connected, (2) connected by a wire, (3) connected by a resistor, or (4) connected by a battery. The goal of the problem-solver is to identify these connections and their values, if applicable.

structure [15]. The task of the problem-solver is to infer the hidden electrical structure by using the tools provided in the simulation interface and collecting data from the four wires protruding from the box (Fig. 1).

The black box problems also have all the characteristics of engineering and scientific problems listed earlier: (1) to solve this problem one needs to use content knowledge about electrical circuits: knowledge about Ohm's law, structural characteristics of circuits, as well as characteristics of different electrical components, (2) the set of required actions to solve the problem is not clear in advance, (3) the problem has multiple parts and multiple possible solution paths. To solve the black box problem, the problem-solver needs to decide on what information to seek, how to get that information, and then how to interpret and use the information that they attain. Therefore, the solution path for the black box problem involves many decisions to be made and different problem-solving practices. At the same time, the number of possible paths was sufficiently constrained that all these possible paths and their associated practices could be characterized. Embedding the problems in an interactive simulation allows the problem-solver to autonomously make all the above-listed decisions.

3.2 Methods and Participants

Experiment 1 – This study was implemented in a project-based upper-level mechanical engineering course at Stanford University. During the course, students completed seven weekly design projects. The first five projects were completed individually, and the last two were completed in groups of two or three students. On Wednesday of each week, participants would receive a project definition sheet as well as reflection prompts to use as they worked on the project. On the following Monday, they would give a 15-minute oral report on their progress to the instructor and receive feedback. On the next Wednesday, they would submit slides and final design files for their project. The course had 17 senior undergraduate or beginning graduate students in mechanical engineering (10 females, 7 males). All except one student was tested using the CCK black box problem at both the start and completion of the course. That student was excluded from the study. Nine students had taken only two courses covering anything about electrical circuits, which we label as the "low experience" group. The other seven students ("high experience" category) had taken multiple courses providing them with extensive backgrounds in working with electrical circuits.

As shown in Fig. 2, the study had three main components: pre-course problem-solving test at the beginning of the quarter, reflection prompts throughout the quarter, and post-course problem-solving test at the end of the quarter.

Pre-test. Students completed the pre-test for problem-solving in the first week of the quarter before they had done any assignments. They first

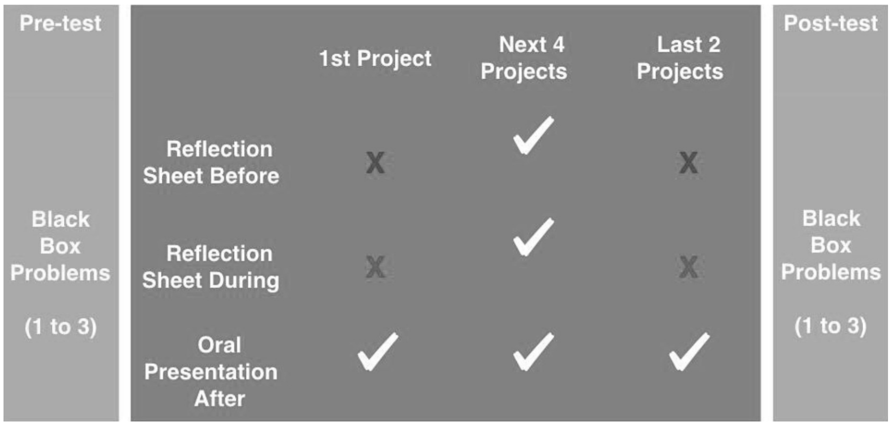

Fig. 2. The design for Experiment 1. The experiment was implemented in a project-based undergraduate mechanical engineering course. The course had seven projects and 17 students, 16 included in the study. Nine students had low experience in electricity and seven students had high experience.

got familiar with the PhET CCK interface and then they solved at least one, and up to three, black box problems, thinking aloud as they did so, spending a total time of 45 minutes. We collected audio and video recordings of the tests, along with the notes that the students took, and a screen-capture video of the simulation.

Reflective prompts. After students had handed in their first project in the second week, the instructor presented the aforementioned problem-solving framework, explained each of the eight practices, and gave examples of a good and bad performance of each practice in the context of a mechanical engineering project. The instructor also emphasized that the practices are generalizable to contexts other than mechanical engineering projects. For the next four projects, students received the project definition along with a worksheet called *Strategizing* (*Appendix A*). They worked on this sheet for the last 10 minutes of the class. The goal of this worksheet was to teach *reflection on problem-definition and assumptions*. In this worksheet, students were explicitly asked to reflect upon how they had defined and decomposed the problem. The worksheet asked students to write down: the definition of the problem; what they already knew that would help them design this project and what they assumed about the problem (*problem definition*); how to break down the project into more manageable sub-projects; and how to budget the time across these sub-projects (*problem decomposition*). Students were encouraged to refer to this worksheet while working on the project.

At the end of the class, students were given a second worksheet named *Reflection* worksheet (*Appendix B*). The goal of this worksheet was to teach reflection on *strategy, knowledge*, and *solution.* In this worksheet, students were asked what strategies they were going to use to carry out the design project, why this strategy was chosen (*reflection on strategy*), what knowledge they already had that would be useful for this project, if they were missing a piece of knowledge how they could get that knowledge (*reflection on knowledge*), how they decided the project was finished, whether the designed product met the constraints of the problem definition, how their design could be tested, whether there was any other possible design, and finally, how they predicted their design would fail (*reflection on solution*). Students were encouraged to refer to and complete this worksheet as they were working on their projects. Students would hand in this worksheet and the strategizing worksheet along with their assignments. Students did not receive any reflection training for the first project, as we wanted the first project to be a preparation for future training [34]. Not having a reflection prompt for the first project versus having it for the next project created contrast and helped students appreciate the important role of reflection in working on an engineering project. As one of the students noted:

> "For [the first project], I found myself getting caught up working on unimportant details or trying to re-CAD components as my design evolved. By the time I got through the initial CAD work, I had run out of time for the most important design considerations (manufacturability, spec'ing components, material choices, etc.) [For the second project and using reflective prompts], by gathering my thoughts and planning the core concepts before attempting to design a solution, I was able to avoid many time-traps such as frequent redesigning of components. This also helped me get more out of the assignment, as I was able to better distribute my time across the various aspects of the project."

The two reflective prompt worksheets were not given out for the last two projects.

Post-test. After finishing all the course requirements, students completed the post-tests. The post-

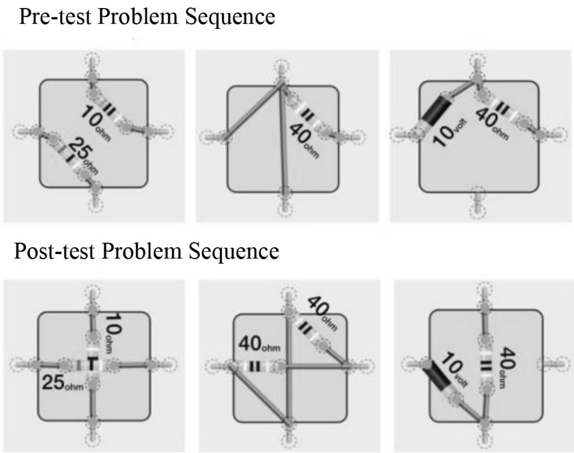

Fig. 3. Problem sequences used for pre- and post-tests of the first experiment. The top row shows the black-box problems used in the pre-test, and the bottom row is the problems used in the post-test. The order of attempting the problems was from left to right, e.g., in the pre-test, students would start with the problem shown on the top left corner, and if time allowed, would proceed to the one shown in the top middle, and then the one on the top right corner. The pre-, post-problems were chosen to be similar but not identical, and the difficulty level increased from the first to the third problem.

tests were conducted in the same way as the pre-tests, but test time was shortened to 30 minutes. Students worked to figure out hidden circuits that were similar, but not identical, to the pre-test circuits. They had similar components, circuit structures, and number of connections but were connected differently as shown in Fig. 3. Students solved at least one and up to three problems in the post-test.

We then analyzed the quality of problem-solving practices and accuracy of final solution on the pre- and post-tests of all the students.

Experiment 2 – The second experiment was carried out in a laboratory setting with 74 students, each participating in a single hour-long interview session. We only recruited low-experience students through a pre-study survey to avoid the ceiling effect observed in study 1. We were able to have both treatment and control conditions in this study to compare the effects of reflective prompts with simple additional practice at the task. Finally, implementing the study in a single experimental session ensured that the students were not learning about electrical circuits outside the study setting. The study design is presented in Fig. 4. Students were randomly assigned to the PR or RP conditions. In both conditions, students first read and signed a consent form, then completed a short tutorial to become familiar with the features of the simulation used for the black box problems and to review basic circuit knowledge (Ohm's law).

In the RP condition, students worked on the first black box problem (black box A), drew their answers to the hidden circuit structure on paper, and then were told whether their solution was correct and shown the correct solution. They then attempted to solve black box problems B and C, without seeing the answer to either. In the PR condition, after doing the tutorial, students worked on black box A, drew their answers on paper, and were then told if their solution was correct. Students then answered two open-ended reflection questions: "what do you think is wrong with your solution?" and "what could you do about it?". Afterwards, they were shown the correct solution for black box A and answered the six reflective prompts shown in Fig. 5. After answering the reflection questions, the PR students attempted black box C.

In both conditions, after finishing black box C,

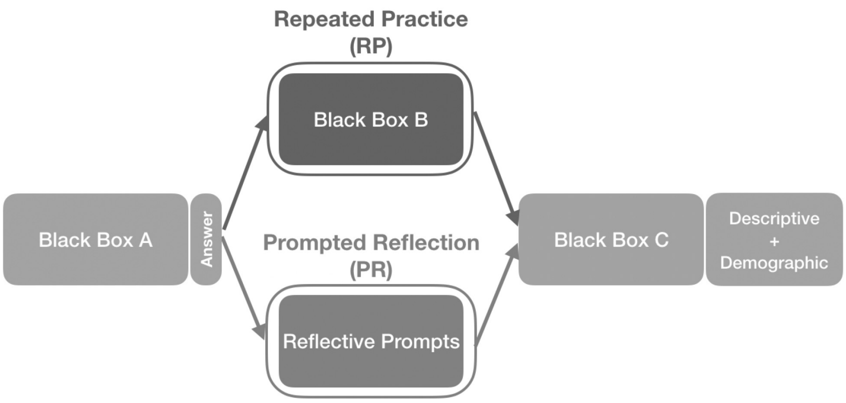

Fig. 4. The design for experiment 2. Students were randomly assigned to the PR or RP conditions. All students at the beginning attempted black box problem A, and then were shown the correct solution. Also, at the end all students answered few demographic and descriptive questions. After seeing the correct solution to black box A, in the Repeated Practice (RP) condition, students attempted to solve black box problems B and C, without seeing the answer to either; in the Prompted reflection (PR) condition, students answered the six reflective prompts, then attempted black box C.

You will have the opportunity to solve a similar black box problem. Given what you have learned from solving the first black box problem, discuss how you can improve upon the following:

Your understanding of the problem

The assumptions you've made about possible solutions

The information you should collect

The approach you have to solving the problem

Keeping track of the information you have about the problem

Confirming that your solution is correct

Fig. 5. The reflective prompts that the students answered after working on the first black box problem in PR condition.

students answered five open-ended descriptive questions about how they solved black box C. They also answered demographic questions about their gender, year in college, major or intended major, and courses taken covering electrical circuits. Students had up to an hour for the study: a maximum of 15 minutes for each black box problem, a maximum of 15 minutes for the open-ended and reflective prompts, and the remaining time spent on the tutorial, descriptive and demographic questions. The Qualtrics online survey platform was used to administer the questions and to collect the student answers. Students were asked to think aloud when working on the black box problems. Audio, video, screen-captured, backend data from the simulation, and notes students took during the study were collected.

Fig. 6 shows black box A, B, and C used for this study. These black box problems were used because they were challenging for these students to solve in the allotted time. This was particularly true for the first problem A because we wanted their struggles with the problem to motivate them to self-reflect and answer the reflective prompt questions.

For the two subsequent problems we used black box B and C, because these two problems are conceptually very similar while structurally not identical. Each hides a battery and a resistor connected to two pairs of wires sharing a common wire.

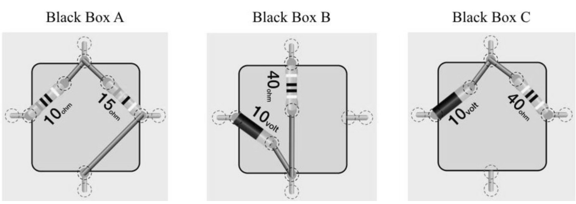

Fig. 6. Black box problems used in study 2. Black box A and C were the first and the last problem attempted by all students. Black box B was only attempted by the students in RP condition, between the first and the last problem.

While students in RP condition did not have the explicit reflective prompts to correct their shortcomings in solving black box A, they had an extra opportunity to practice with a problem quite similar to black box C. Also, the similarity of these two problems allowed us to compare the students' performances on the two problems.

To recruit participants, we sent out an online form to different Stanford students' mailing lists. In this form, students were asked about their major as well as the courses they had taken in high school or college that covered electrical circuits. From the students who filled the forms, we recruited 74 students (43 females, 31 males) with some, but not extensive, background in electrical circuits. This included students who had taken a high school or university physics course covering simple electrical circuits but not majoring in electrical engineering or physics, as those majors would give them more exposure to circuits. We excluded four of the participants: one of them had surgery less than 24 hours before the experiment, the other one (despite their answer to the recruitment form) had no electricity background, one had to leave before finishing the experiment, and one had already participated in one of our other studies with black box problems. The remaining 70 students were rated by their level of experience with electrical circuits: students who had taken *high school physics* course covering electrical circuits (HP), students who had taken *honors or advanced placement* high school physics courses covering electrical circuits (HAP), and students who had taken introductory *undergraduate* courses covering electrical circuits (UN). The number of students in each level across the two conditions is shown in Table 1.

To analyze students' problem-solving, we used the audio, video as well as screen-captured data. To analyze their final solutions, we used the solutions students drew on paper. We used similar but not

Table 1. Number of students in each experience level across conditions

	HP	HAP	UN
RP	6	17	12
PR	5	15	15

identical procedures for measuring the quality of problem-solving practices demonstrated by the subjects in the two experiments.

3.3 Practice and Problem-Solving Score

Experiment 1 – Students received a score from 0 to 3 for seven of the eight problem-solving practices, depending on how effectively they performed that practice in each black-box problem they attempted to solve. We did not score them on data-interpretation because we have seen that performance of this practice is dominated by their experience with the content. The practices' scores were added together to define the problem-solving score for each student. Students' problem-solving score is the sum of students' scores for these seven practices, ranging from 0 to 21. Finally, each student's scores of specific practices as well as their overall problem-solving scores were averaged for the number of problems attempted during the pre- or post-test. The details of the scoring rubric for each of the individual practices is discussed in [13].

Experiment 2 – The sample size of this study was significantly larger than the previous studies, and we had to make the video coding scalable to analyze the data of 70 students. Therefore, we shortened and adjusted the coding scheme to focus on the practices that were directly addressed by the prompted reflection questions in our study. We coded for the following practices: problem definition, problem decomposition, data collection, data recording, and solution reflection. For each attempted problem, we coded students' performances in each of these practices. Depending on their performance, the student received a score of 0 to 3. Students' problem-solving score is the sum of students' scores for these five practices, ranging from 0 to 15.

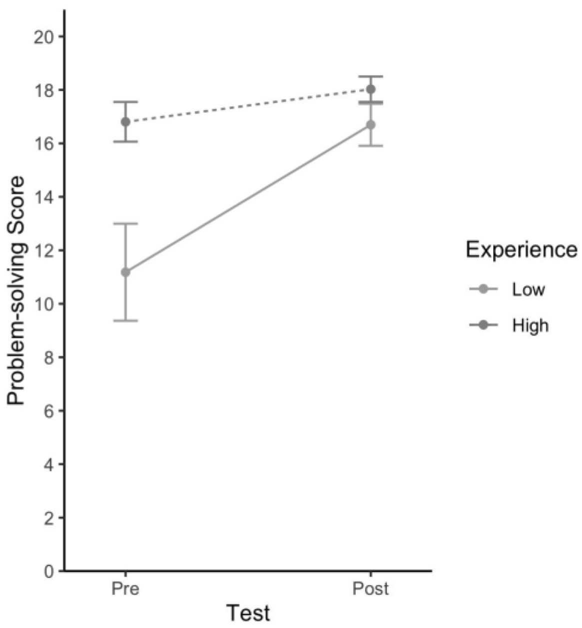

Fig. 7. Problem-solving scores across experience levels and the pre- and post-tests. Panels correspond to the experience level. The experience listed here is based only on what courses they have taken covering the electricity content. Low experience level is shown by the blue solid line and high experience level is shown by the green dashed line.

4. Results

4.1 Experiment 1: Pre- to Post-Test Change in Problem-Solving Practices Scores (RQ1 & RQ2)

Fig. 4 shows the progression of average problem-solving scores from pre- to post-tests for the high and low experience level groups. The scores increased for both levels, and the increase was larger for the low-experience students, probably because of the ceiling effect for the high experience group. To analyze the statistical significance of the improvement seen in Fig. 7, we ran a mixed model repeated measure regression of the problem-solving practices score on the fixed effects of test and experience level, while controlling for the random effect of students. The analysis showed that the problem-solving score of all students increased from pre- to post-test. Low experience students increased by an average of 5.52 points (1.31 S.D.) (p = 0.0007). The increase for high experience students was 1.23 points (0.2 S.D.) (p = 0.044). In the post-test, the high experience group achieved scores that were very close to the maximum score attainable, so the measured improvement in the test can be considered as a lower limit of their actual improvements.

4.2 Experiment 1: Improvement of Specific Problem-Solving Practices (RQ1)

Fig. 8 shows the changes in specific problem-solving practices' scores between pre and post-tests. Scores mostly increased. This was especially the case for low-experience students who had the most room to improve. To statistically verify the patterns observed for individual practices, we used a mixed-model repeated measure regression to predict the score of each practice based on experience level, tests (pre- versus post-test), and the interaction term. After adjusting for Bonferroni correction for seven practices, all changes are statistically significant at the $p < 0.05$ level, except for reflection on problem-definition and reflection on knowledge.

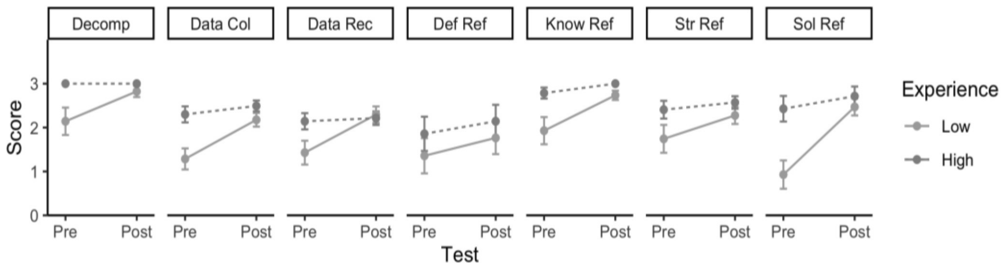

Fig. 8. Problem-solving scores across experience levels, practices, and pre- and post-tests. From left to right the practices indicate: problem definition and decomposition, data collection, data recording, reflection on problem definition and assumptions, reflection on knowledge, reflection on strategy, and reflection on solution (verification of solution). When a participant had attempted more than one problem in either test, the score for each practice would be averaged across the attempted problems. Low experience level is shown by the blue solid line and high experience level is shown by the green dashed line.

Reflection on solution showed the most pronounced improvement across both experience levels, with an average improvement of 1.38 points from the pre- to post-tests (p_{adjusted} = 0.007). This is likely because reflection on solution is the least complex and hence most straightforward reflection practice to improve. This may also be due to more training on this practice. For each project, the instructor asked, "If and how would your design break (fail)." This question is a manifestation of reflection on solution. In the post-test, after arriving at a tentative solution, several students said, "let me see how I can break my solution."

4.3 Experiment 1: Percentage of Correct Solutions (RQ3)

There was a substantial variation in how many hidden circuits were correctly determined by each participant in pre- and post-tests. However, everyone attempted at least the first problem. Fig. 9 shows the percentage of correct solutions for the first problem across experience levels and tests. The mixed-effect repeated measure logistic regression was used to analyze the probability of solving the problem correctly based on test (pre- or post-test) and experience level, while controlling for the random effect of students. The high-experience students performed better on both pre- and post-tests. Overall, the probability of students correctly solving the first problem increased from the pre- to post-tests (p = 0.028). This increase was not significantly different across experience levels and problems. The results were similar when we analyzed the solutions to the first two problems attempted in each interview. So, students not only improved in their problem-solving practices, but

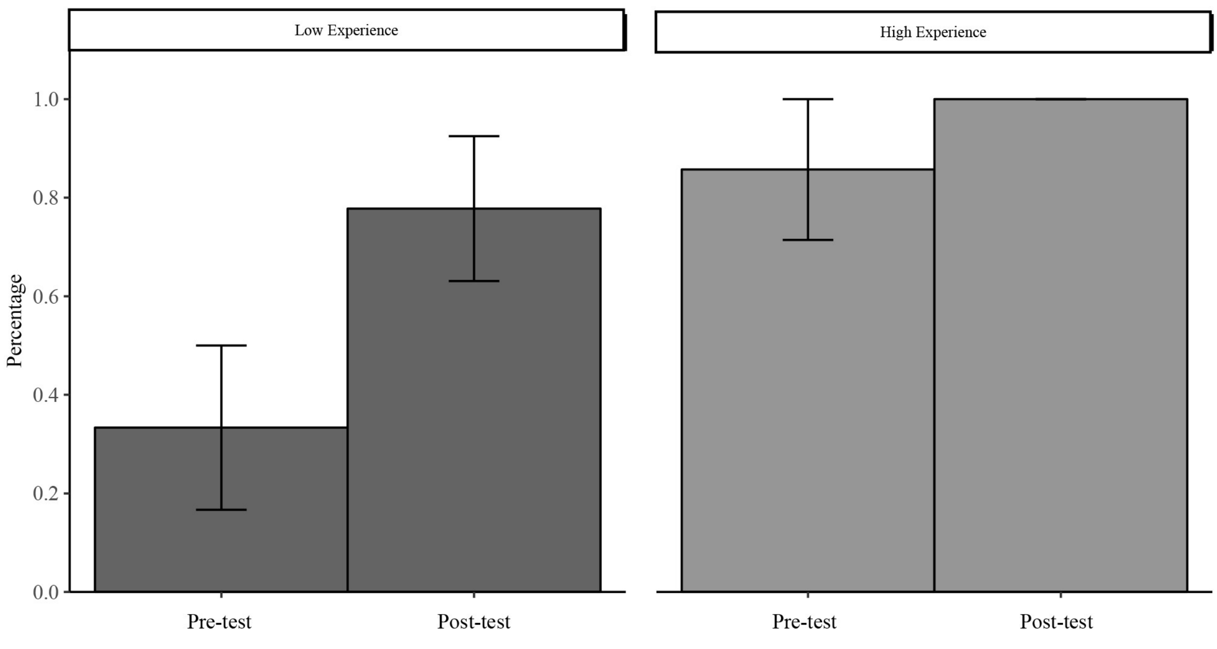

Fig. 9. Percentage of the correct solution for the first attempted problem across different experience levels, and pre- and post-tests. Note that high-experience students all achieved the correct solution for the first problem of the post-test, so there is a ceiling effect. They might have shown more improvement had we used a harder problem.

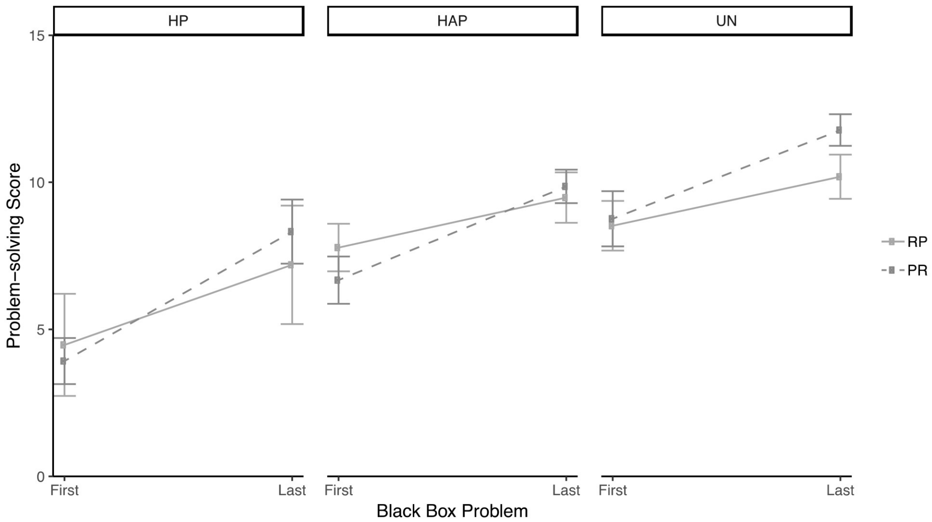

Fig. 10. Progression of the problem-solving score from the first(A) to the last(C) problem across conditions and levels of experience. The maximum problem-solving score is 15, and the error bars represent standard errors. Repeated Practice (RP) is shown by the blue solid line, and Prompted Reflection (PR) is shown by the green dashed line.

this improvement also carried over to their solution of the problem.

4.4 Experiment 2: Change in Problem-solving Practices Scores (RQ1)

First to last problem comparison. Fig. 10 shows the progression of the problem-solving practices scores from the first to the last black box problem across conditions and experience levels. It can be seen that while both conditions improved, the slopes were steeper for the PR condition across all experience levels.

We used multivariable linear mixed-model repeated measure regressions to analyze the statistical significance of the differences in the progression of students' problem-solving scores across conditions and experience levels. To find the simplest best fitting model for the data, we started from the basic additive model with the main effects of condition, experience level, and black box problem (first (A) or last (C)), and the random effect of students. To this basic model, we added interaction terms between main effects, one by one, to test whether any of these interaction terms would improve the model fit significantly. The model fits were compared using chi-square tests. Adding the interaction between problem and condition significantly improved the model fit ($\chi2(1) = 5.77$, $p = 0.016$). No other interaction term further improved the fit; and the fit of the most complex model with all the interactions between problem, condition, and experience level was not significantly different from the model with only the interaction between problem and condition ($\chi2(6) = 2.99$, $p = 0.810$). Therefore, the model with problem, condition, experience level, and the interaction between problem and condition is the simplest best fitting model to describe the data. Table 2 presents the results of this regression model.

The regression analysis shows that students across the two conditions started similarly and their problem-solving scores for the first black box was not different ($p = 0.515$). While both conditions showed improvement in problem-solving, this improvement was about two-fold higher for prompted reflection (PR) compared to repeated

Table 2. The results of regression analysis for the progression of problem-solving score from the first to the last black box problem. The results are based on the simplest best-fitting regression model. The problem-solving score is normalized. The baseline for the condition variable is set to RP, and the experience level is set to HP. The numbers in parentheses present standard errors of regression coefficients. The most relevant term is the last one, which shows the difference in improvement between the two conditions.

Predictors	**Progression of Normalized Problem-solving Score (first to last)**
(Intercept)	b = –1.04 (0.26) $p = 0.0002$
Condition (RP = 0, PR = 1)	b = –0.14 (0.21) $p = 0.52$
Problem (first = 0, last = 1)	**b = 0.52 (0.12)** $p < 0.0001$
HAP Experience Level	**b = 0.70 (0.28)** $p = 0.01$
UN Experience Level	**b = 1.09 (0.29)** $p = 0.0003$
Condition * Problem	**b = 0.40 (0.17)** $p = 0.02$
	Observations = 140 AIC = 335.57

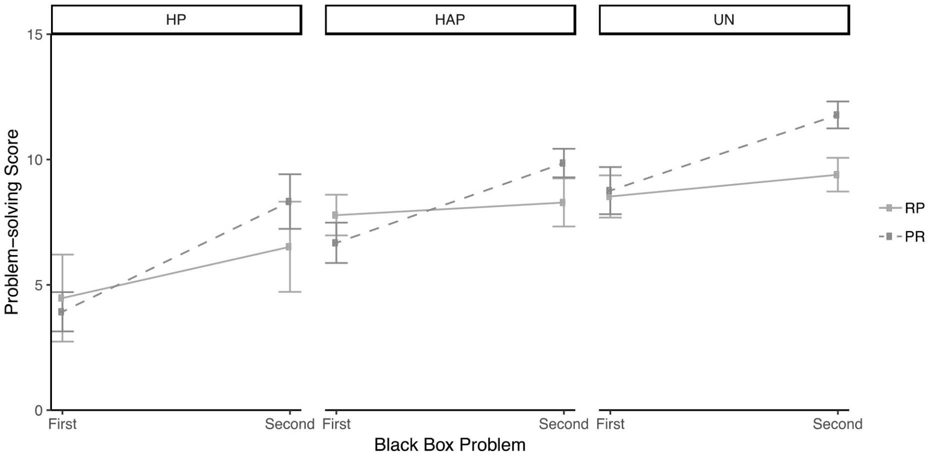

Fig. 11. Progression of the problem-solving score from the first(A) to the second (B/C) problem across conditions and levels of experience. The maximum problem-solving score is 15, and the error bars represent standard errors. Repeated Practice (RP) is shown by the blue solid line, and Prompted Reflection (PR) is shown by the green dashed line.

practice (RP) condition. The average problem-solving score in PR condition improved by 0.92 standard deviation ($p = 0.018$) from the first to last problem, while this improvement was 0.52 standard deviation for RP condition ($p < 0.0001$). While the initial scores varied with student experience level, the improvements were comparable across all levels. These results indicate that giving students specific reflection prompts about their problem-solving led to substantially greater improvement in problem-solving compared to simply providing them with opportunities for repeated practice of problem-solving.

First to second problem comparison. The results above were based on the first to last problem and comparing more opportunities to practice in the RP group with time for reflection prompts in PR group (three problems vs. two problems plus prompts). Alternatively, we can compare the effect of prompts versus no prompts for the same amount of practice by comparing the problem-solving improvement from first to second problems across conditions. The analysis is the same as above for the PR group, except for RP group it is analyzing improvement from problem A to problem B, instead of to problem C in the previous analysis. By design, problems B and C were very similar and equally difficult. Fig. 11 shows the problem-solving scores from the first to the second black box problem across conditions and experience levels.

Table 3 presents the results of regression analysis for the problem-solving progression from the first to the second problem across conditions. The simplest best fitting model of this analysis was the same as in the previous analysis. However, this time, the average problem-solving score of students in the RP condition increased 0.25 standard deviation ($p = 0.040$) from the first to the second problem (A-B), which is about one quarter as large as the 0.92 standard deviation improvement (A-C) seen for the PR students ($p = 0.0002$).

Table 4 summarizes the progression of problem-solving scores across conditions. The PR problem-

Table 3. The results of regression analysis for the progression of problem-solving score from the first to the second black box problem. The results are based on the simplest best-fitting regression model. The problem-solving score is normalized. The baseline for the condition variable is set to RP and the experience level is set to HP

Predictors	Progression of Normalized Problem-solving Score (1st to 2nd)
(Intercept)	b = –0.95 (0.26) $p = 0.0006$
Condition (RP = 0, PR = 1)	b = –0.13 (0.21) $p = 0.51$
Problem (first = 0, second = 1)	**b = 0.25 (0.12)** $p = 0.04$
HAP Experience Level	**b = 0.67 (0.28)** $p = 0.02$
UN Experience Level	**b = 1.08 (0.29)** $p = 0.0003$
Condition * Problem	**b = 0.67 (0.17)** $p = 0.0002$
	Observations = 140 AIC = 339.97

Table 4. Summary of the progression in problem-solving scores for prompted-reflection (PR) and extra repeated-practice (RP) in units of standard deviations

	Progression of problem-solving score	
	RP	PR
1st to last problem	0.52	0.92
1st to 2nd problem	0.25	0.92

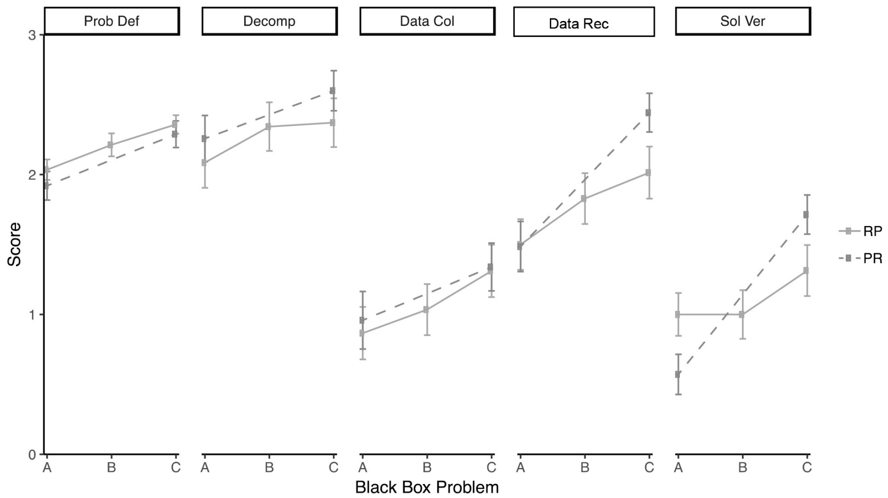

Fig. 12. Progression of problem-solving practices from the first to the last problem across conditions. The maximum score for each practice is 3, and the error bars represent standard errors. From left to right, labels indicate: problem definition, problem decomposition, data collection, data recording, and solution reflection. Repeated Practice (RP) is shown by the blue solid line, and Prompted Reflection (PR) is shown by the green dashed line

solving condition improved about twice as much as the RP condition when comparing first to last problems. When comparing results for the same number of practiced problems (first to second problem comparison), students in the PR condition improved nearly four times more than students in the RP condition.

4.5 Experiment 2: Improvement in Specific Practices (RQ1)

Fig. 12 shows the progression of different problem-solving practices across conditions and levels of experience. The practices of data recording and solution reflection showed the greatest improvement. This is consistent with the results from study 1. To statistically verify the patterns described above, we used a mixed-model repeated measure regression to predict the score of each practice based on experience level, black box problem (first versus last), condition, and the interaction between condition and problem. The improvement for the data recording is significant for both conditions ($p_{\text{adjusted}} = 0.003$) and the improvement is larger for the PR condition. The improvement in solution reflection, while not significant for the RP condition ($p_{\text{adjusted}} = 0.22$), is significant for PR condition at the $p < 0.05$ level and remains significant after multiple comparison correction ($p_{\text{adjusted}} = 0.002$).

The average data collection score is low across problems and conditions. This is not surprising as participants in this study have only limited background knowledge and experience in circuits. Deciding on what data to collect from the black box and how to effectively collect them is strongly influenced by experience. If one has very limited experience with voltmeters, it would be difficult to use it to appropriately to measure suitable voltage differences and resistances between wires. Data recording and reflection on solution are less experience-dependent practices and showed the most improvement in these two studies. A simple prompt to think about how to keep better track of their data and to think how to confirm their solution is enough to have students to engage in these practices effectively.

4.6 Experiment 2: Progression of Correct Solutions (RQ3)

This study was mainly focused on teaching a subset of problem-solving practices. While we examined and reported the number of students across conditions who had the correct solution for each problem, the result is secondary in this study. Although students' problem-solving practices improved, the improvement in the probability of obtaining the

Table 5. The distribution of students who correctly solved black box problems across conditions

	Incorrect	Correct
Black box A (First black box)		
RP	34	1
PR	34	1
Black box C (Last black box)		
RP	29	6
PR	31	4
2nd Black box		
RP	33	2
PR	31	4

Table 6. Students' reply to the debriefing questions of experiment 2

	No	Maybe	Yes
Did you find the reflection questions helpful?	%18	%3	%79
Was there any added value in answering reflection questions?	%19	%3	%78

correct solution was not large, primarily due to the limited time available. We gave participants a fixed amount of time to work on each problem for a number of practical reasons. This time duration was chosen to be long enough to evaluate their problem-solving practices but not sufficiently long for most students to solve the problems. A student's success at achieving a correct solution depends on having sufficient time for their level of experience. Unlike in study 1, where students had sufficient time for a substantial fraction of them to solve the problem, here relatively few did, and hence their improved practices had little impact on the probability of producing a correct solution as shown in Table 5.

4.7 Experiment 2: Student Perception of Reflective Prompts (RQ4)

Perceived helpfulness of the reflective prompts. At the end of each session, we asked the students in the PR condition: "Did you find the questions you answered between black box [A] and [C] helpful at all?" and "Do you think there was any added value in answering those questions besides seeing the black box [A] answer?" Table 6 shows students' answers to these questions.

As shown in Table 6, a significant majority of students thought the reflective prompts were helpful ($p < 0.0001$). As one student replied:

> "Reflection questions helped me realize how narrow-minded my approach was, and I should use more of the tools provided. Forcing to reflect is helpful; and if I just saw the answer, I would just say okay, I am wrong, and I would not try many different possibilities without answering the questions. I had so much difficulty in high school physics and answering the reflective questions would have helped me."

Similarly, an overwhelming majority of students believed there was added values in answering the reflection prompts ($p = 0.0007$). As one student said:

> "Seeing the right answer was not enough because I am more interested in the process, and the reflection questions helped me realize how the right answer was achieved."

Or as another student mentioned,

> "There was some added value. Once I saw the solution, I realized immediately what I was doing wrong, but the questions served more so as thinking further about how I can change my problem-solving approach for the next [problem], and test and use voltmeter and ammeter."

Overall, students believed the reflective prompts were helpful and there were added values in answering them. They explained that either the prompts helped them think more in depth about the process of solving the problem or writing down the answers helped them better commit to their approach for the next problem.

Problem-solving practices reflected vs. improved. Of the 30 students who answered both debriefing questions, 24 mentioned recognizing their shortcomings in some specific problem-solving practice(s) and said that the reflection prompts helped them realize they needed to improve in these practices. An average of 2.1 practices per student was mentioned. For each of these 24 students, we examined their scores on the practices they mentioned. From the first problem to the last, they showed improvement in the scores on the practices they mentioned 84 percent of the time; no change 6 percent of the time, and a decline 10 percent of the time. This result indicates that nearly all students successfully acted on the reflective prompts that revealed their weaknesses to them.

Overall, the analysis of students' responses to the debriefing questions suggests that students found the reflective prompts helpful and were mostly successful in improving the specific practices that they noted in their reflections.

5. Discussion

In the two experiments presented in this study, we examined whether students can be taught to be better problem-solvers by engaging them in reflective practices via prompts. The two experiments together show that prompting students to reflect would improve student overall problem-solving. Prompting students to reflect would benefit students different problem-solving, particularly their reflection on solution and data recording, as these practices are less content dependent and simply prompting students to reflect upon them can lead to improvement. On the other hand, problem-definition, decomposition and data collection practices showed smaller improvement, as the effectiveness of engaging in these practices is content dependent, and mere reflection on these practices would not be enough to sufficiently improve. These experiments together show that improvement gained from prompted reflection can transfer across contexts and outweighs the benefit of repeated practice of similar problems.

In the first experiment, the reflective prompts were implemented in the context of working on mechanical engineering course projects. Before and after the course, students attempted similar electrical black box problems to evaluate their problem-solving. After the course, students had significantly higher problem-solving scores and were more likely to solve the problem correctly. All students significantly improved after the course. These improvements show that not only can problem-solving be taught, but that the reflective practices transcend across disciplinary boundaries. Both of these results are notable. These results underscore that we can train students to become more reflective problem-solvers through coursework. Furthermore, such training would have a broad educational impact as students can transfer such skills to different contexts. This finding suggests that these reflective practices and the benefit of prompting and practicing them during instruction are generalizable across very different contexts and can be used to teach problem-solving practices.

The second experiment was a follow up experiment to check the effects of repeated practice on the results of experiment 1, and make sure the observed improvement in students' problem-solving was because of reflective prompts, not mere repeated practice of similar problems in pre- and post-test. This second experiment compared the effect of receiving reflective prompts with repeated practice of similar problems. Within one hour of lab experiment, one condition solved two black problems, while receiving reflective prompts in between the first and second problems to consider improvement in their problem-solving. The other condition attempted three black box problems without any reflection prompts in between. The result of this experiment showed that students who received reflection prompts on specific practices acted on and improved the practices that they were able to note in their own reflections. This improvement was substantially greater than the effect of simply having more practice in doing problems. This result also confirms that the pre- to post-course improvement observed in the first experiment was mainly due to the reflective prompts resulting in improved problem-solving practices rather than other factors such as repeated practice of solving black box problems in pre- and post-test. By repeatedly practicing different instances of the same problem, people are capable of reflecting on their own problem-solving practices and thereby improving without any explicit training. However, this benefit of repeated practice is less than the benefit provided by prompting student to reflect on specific aspects of their problem-solving during problem solving process. We conjecture that most students would be unable to adequately identify and address their shortcomings in the absence of such prompts because of the complexity of problem-solving process [35].

This finding has implications for instruction. It highlights that teaching students to become better problem-solvers, particularly with regard to reflective practices, is best done by explicitly scaffolding problem-solving practices, rather than just having students practice solving problems. The prompts used in the two studies focused on reflective practices with a relatively light-touch approach. These prompts only asked students to engage in different reflective practices to identify any possible improvements in their problem-solving. This instruction made students more intentional about the decisions they made in their problem-solving, but it did not guarantee students would make better decisions and use more effective practices. Making better decisions requires reflection on the decisions to be made, knowing the alternative choices for those decisions, and recognizing how to choose the best choice. As we saw in both experiments, for some practices like data collection, only receiving reflective prompts did not lead to much of an improvement. As for effective data collection practice, one needs content knowledge about the problem to evaluate different alternative choices. Therefore, just as providing more opportunities to solve problems without explicit problem-solving training would not be sufficient to improve problem-solving, neither would teaching problem-solving practices in a content-free and abstract fashion. The intertwined nature of content knowledge and problem-solving practices requires students to acquire content knowledge as well as understand how this knowledge manifests itself in better choices for problem-solving practices. This can be achieved by providing students more just-in-time feedback about the decisions they make, and the knowledge required to make such decisions.

6. Limitations

The study presented here has a number of limitations that should be addressed in future works. First, while we helped students to identify what to reflect upon by giving them specific reflective prompts, and in the second study we also scaffolded when to reflect by providing the prompts between the first and the second problem; we did not train students on how to reflect. In other words, we did not provide students feedback on their answers to reflective prompts. Second, we did not provide students with content knowledge that was needed in some cases for more effective reflection. For example, we did not provide students with required electricity content knowledge that could have helped them with more effective reflection upon data collection strategies for solving the black box problems. Future studies should expand on this work by further examining the effect of not only prompting students to reflect, but also providing feedback on the quality of their reflection as well as providing further required knowledge to improve their reflection.

7. Conclusion

The results of this study show that students found being prompted to reflect helpful for their problem-solving and these reflective prompts improve their problem-solving in science and engineering domains. The reflective prompts encourage student to evaluate their problem-solving processes and consider how to improve them. The improvement transfers across contexts, as in the first experiment students received reflective prompts during their four mechanical engineering course projects, and they showed improvement in solving electric circuit problems from pre- to post-course test. The benefit of reflective prompts also outweighs repeated practices of similar problems, as shown in the second experiment, the students who received reflective prompts improved more in their problem-solving compared to the ones who attempted more similar problems but without any reflective prompts.

References

1. National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*, National Academies Press, Washington, DC, 2012.
2. NGSS Lead States, *Next Generation Science Standards: For States, By States*, National Academies Press, Washington, DC, 2013.
3. H. J. Passow and C. H. Passow, What competencies should undergraduate engineering programs emphasize?, *A Systematic Review. Journal of Engineering Education*, **106**(3), pp. 475–526, 2017.
4. S. Sheppard, A. Colby, K. Macatangay and W. Sullivan, What is engineering practice?, *International Journal of Engineering Education*, **22**(3), pp. 429–438, 2006.
5. Kirn and L. Benson, Engineering students' perceptions of problem solving and their future, *Journal of Engineering Education*, **107**(1), pp. 87–112, 2018
6. D. Dailey and A. Cotabish, E is for engineering education, Cultivating applied science understandings and problem-solving abilities, in B. Macfarlane (ed), *STEM Education for High-Ability Learners*, Routledge, New York, pp. 71–83, 2021.
7. ABET Engineering Accreditation Commission. Criteria for accrediting engineering programs, Baltimore, MD, 2022.
8. E. Burkholder and C. Wieman, Impact of decision-making in capstone design courses on students' ability to solve authentic problems, *International Journal of Engineering Education*, **37**(3), pp. 650–662, 2021.
9. P. Cuevas, O. Lee, J. Hart and R. Deaktor, Improving science inquiry with elementary students of diverse backgrounds, *Journal of Research in Science Teaching*, **42**(3), pp. 337–357, 2005.
10. B. Memarian and S. Mccahan, Constructive alignment integrated rating (CAIR) in the assessment of engineering problem-solving, *International Journal of Engineering Education*, **37**(5), pp. 1441–1453, 2021.
11. K. Duncker and L. S. Lees, On problem-solving, *Psychological Monographs*, **58**(5), i, 1945.
12. A. Newell and H. A. Simon, *Human problem solving (Vol. 104, No. 9)*, Prentice-Hall, Englewood Cliffs, NJ, 1972.
13. G. Polya, *How to solve it: A new aspect of mathematical method (No. 246)*, Princeton University Press, Princeton, NJ, 2004.
14. F. Reif, *Understanding Basic Mechanics*, John Wiley & Sons, Hoboken, NJ, p. 288, 1995.
15. Authors, *Improving problem-solving through reflection*, Stanford University, 2018.
16. A. Price, C. J. Kim, E.W. Burkholder, A.V. Fritz and C. E. Wieman, A detailed characterization of the expert problem-solving process in science and engineering: Guidance for teaching and assessment, *CBE – Life Sciences Education,* **20**(3), ar43, 2021.
17. M. F. Rubinstein, *Patterns of Problem Solving*, Prentice-Hall, Englewood Cliffs, NJ, pp. 8–10, 1975
18. C. J. Atman, R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg and J. Saleem, Engineering design processes: A comparison of students and expert practitioners, *Journal of Engineering Education*, **96**(4), pp. 359–379, 2007.
19. P. S. Steif, J. M. Lobue, L. B. Kara and A. L. Fay, Improving problem solving performance by inducing talk about salient problem features, *Journal of Engineering Education*, **99**(2), pp. 135–142, 2010.
20. K. B. Wendell, C. G. Wright and P. Paugh, Reflective decision-making in elementary students' engineering design, *Journal of Engineering Education*, **106**(3), pp. 356–397, 2017.
21. J. A. Turns, B. Sattler, K. Yasuhara, J. Borgford-Parnell and C. J. Atman, Integrating reflection into engineering education, *In Proceedings of the ASEE Annual Conference and Exposition*, ACM, New York, NY, **35**, p. 64, 2014.
22. S. A. Ambrose, Undergraduate engineering curriculum: The ultimate design challenge, *The Bridge*, **43**(2), pp. 16–23, 2013.
23. G. Salomon and T. G. E. Globerson, The computer as a zone of proximal development: internalizing reading related metacognition from reading partner, *Journal of Educational Psychology*, **81**(4), 1989.
24. J. W. Coffey and B. Owsnicki, Introducing a reflective activity into the design process in an advanced computer programming course, *Journal of Computing Sciences in Colleges*, **31**(5), pp. 29–37, 2016.
25. J. W. Coffey, A study of the use of a reflective activity to improve students' software design capabilities, *SIGCSE '17: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, Seattle, WA, pp. 129–134, March 2017.
26. S. E. Volet, Modelling and coaching of relevant metacognitive strategies for enhancing university students' learning, *Learning and Instruction*, **1**(4), pp. 319–336, 1991.
27. B. Kramarski and Z. R. Mevarech, Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training, *American Educational Research Journal*, **40**(1), pp. 281–310, 2003.

28. A. H. Schoenfeld, Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics, in D.A. Grows (eds), *Handbook of Research on Mathematics Teaching and Learning*, Marmillan, New York, NY, pp. 334–370, 1992.
29. J. H. Hanson and J. M. Williams, Using writing assignments to improve self-assessment and communication skills in an engineering statics course, *Journal of Engineering Education*, **97**(4), pp. 515–529, 2008.
30. R. Bailey, Exploring design process learning through two reflective prompts, *The International Journal of Engineering Education*, **36**(2), pp. 568–573, 2020.
31. C. Cardoso, A. Hurst and O. G. Nespoli, Reflective inquiry in design reviews: The role of question-asking during exchanges of peer feedback. *The International Journal of Engineering Education*, **36**(2), pp. 614–622, 2020.
32. K. Perkins, W. Adams, M. Dubson, N. Finkelstein, S. Reid, C. Wieman and R. LeMaster, PhET: Interactive simulations for teaching and learning physic, *The Physics Teacher*, **44**(1), pp. 18–23. 2006.
33. Circuit Construction Kit: DC, accessed February 2, 2023, https://phet.colorado.edu/en/simulations/circuit-construction-kit-dc.
34. J. D. Bransford and D. L. Schwartz, Chapter 3: Rethinking transfer: A simple proposal with multiple implications, *Review of Research in Education*, **24**(1), pp. 61–100, 1999.
35. J. Prior, S. Ferguson and J. Leaney, Reflection is hard: teaching and learning reflective practice in a software studio, *In Proceedings of the Australasian Computer Science Week Multiconference*, ACM, New York, NY, February 1, 2016

Shima Salehi is an Assistant Professor at Stanford Graduate School of Education, and the director of IDEAL research lab, the research component of Stanford IDEAL initiative to promote *inclusivity, diversity, equity, and access in learning communities*. Her research focuses on how to use education and cognitive theories to improve effectiveness and equity in STEM education. For effective science and engineering education, Dr. Salehi has studied effective scientific problem-solving and developed empirical framework for main problem-solving practices to train students in. Based on these findings, she has designed instructional activities to provide students with explicit opportunities to learn these problem-solving practices. These activities have been implemented in different science and engineering courses. For Inclusive science and engineering, she examines different barriers for equity in STEM education and through what instructional and/or institutional changes they can be addressed.

Karen Wang is a PhD candidate at the Stanford Graduate School of Education. Her research focuses on defining, measuring, and teaching complex problem-solving skills using data-enriched assessments and learning analytics techniques. Previously, she worked as a technology teacher in the nation's first Chinese-English bilingual immersion school in San Francisco and a distance educator at the Johns Hopkins Center for Talented Youth in Baltimore. She holds a BA in College of Social Studies and Economics from Wesleyan University and an MA in Learning, Design and Technology from Stanford University.

Michael Flynn, Michael's research focuses on the development and assessment of expertise in mechanical engineering design. He teaches courses in mechanical engineering (ME181 and ME324). He received his PhD in Mechanical Engineering and Masters in Education from Stanford University.

Carl E. Wieman holds a joint appointment as Professor of Physics and of the Graduate School of Education. He has done extensive experimental research in atomic and optical physics. His current intellectual focus is now on undergraduate physics and science education. He has pioneered the use of experimental techniques to evaluate the effectiveness of various teaching strategies for physics and other sciences, and served as Associate Director for Science in the White House Office of Science and Technology Policy.