The Effect of Providing Motivational Support in Parsons Puzzle Tutors

Amruth N. Kumar
Ramapo College of New Jersey, Mahwah, USA
amruth@ramapo.edu

Abstract. In response to student feedback on a tutor on Parsons puzzles on the programming concept of sequence, we incorporated three features meant to improve the motivation of the student solving the puzzles. We compared the performance of students before and after introducing these features. We found that introduction of motivational supports did not affect pre-post improvement, and therefore, the amount of learning. Students who were provided motivational supports spent more time per puzzle than those who were not.

Keywords: Parsons puzzle, Programming, Tutor, Evaluation.

1 Introduction

Parsons puzzles have been gaining popularity as a mechanism for teaching programming concepts. In a Parsons puzzle [1], the student is presented a problem statement, and the program written for it. The lines in the program are provided in random order. The student must re-assemble the lines in their correct order.

Epplets (epplets.org) are a suite of tutors that we developed to help students learn programming concepts in C++/Java by solving Parsons puzzles. Each tutor in the suite presents puzzles to the student, has the student solve the puzzles, and provides feedback if the solution is incorrect.

We first deployed the tutors in fall 2015. They were used by students in the introductory programming course from six different undergraduate institutions in fall 2015 and spring 2016. The feedback provided by this user group included that they needed better motivational supports when solving puzzles with the tutors. In particular, they wanted the tutors to provide them a better sense of progressing through the material. In response, we incorporated the following features into the tutors before re-deploying them in fall 2016:

1. When each puzzle was first presented, the tutor listed the number of lines in the puzzle and therefore, the number of drag-and-drop actions with which students should aim to solve the puzzle;
2. In addition to summarizing each drag-and-drop action taken by the student as feedback, the tutor also enumerated the action, e.g., the enumeration 3 in:
   3. “Moved from problem to solution at line 8: double loan;”
So, students could keep track of the number of actions they had already taken and the number of actions remaining to solve the puzzle;

3. After the student had solved each puzzle, the tutor displayed the progress of the student, i.e., the number of puzzles the student had solved.

These features track progress, one of the three components of gamification, and thereby provide extrinsic motivation [2]. We evaluated the effect of providing these features on the learning of students by comparing the data collected in fall 2015-spring 2016 (control group before the features were introduced) with that collected in fall 2016 (experimental group after the features were introduced). The tutor we used for this study was on the programming concept of sequence.

The Protocol: The tutor administered pre-test-practice-post-test protocol:

- **Pre-test:** This consisted of one puzzle. If the student solved the puzzle correctly, the tutor ended the session. If the student solved the puzzle partially or incorrectly, the tutor scheduled additional puzzles as practice.

- **Adaptive Practice:** The tutor presented additional puzzles until the student had mastered the concept of sequence.

- **Adaptive Post-test:** After the student had demonstrated mastery during practice, the tutor presented a post-test puzzle. If the student solved it correctly, the session was terminated. Otherwise, the student was returned to solving additional practice puzzles, followed by a repeat post-test.

The entire protocol was limited to 30 minutes. It was administered by the tutor seamlessly, back-to-back with no breaks in between the three stages.

The Subjects: The subjects were students in the introductory programming course. In fall 2015 and spring 2016, the tutor was used by students from 6 undergraduate institutions. In fall 2016, it was used by students from 12 institutions – both high schools and undergraduate institutions. The students used the tutor over the web as after-class assignment, on their own time.

The Design: We considered the following dependent variables:

1. The number of puzzles solved: students solved only one puzzle during the pre-test and post-test, but solved multiple puzzles during adaptive practice.

2. The average score per puzzle solved during pre-test, practice and post-test. Each puzzle has only one correct solution. A puzzle containing n lines of code can be solved with n drag-and-drop actions. Allowing for unintentional mistakes, a student who solved a puzzle with 1.1n or fewer actions was given full credit. Thereafter, partial credit was awarded inversely proportional to the number of unnecessary actions taken by the student. The normalized score on each puzzle, calculated by dividing the number of student actions (after negative grading) by the number of lines in the puzzle, ranged from 0 to 1.0 per puzzle.

3. The average time spent per puzzle during pre-test, practice and post-test.

2 Data Analysis and Results

Univariate ANOVA analysis of the score on the pre-test puzzle showed no significant difference between control and experimental groups \(F(1,186) = 2.883, p = \)
0.091]. So, the two groups were comparable. However, ANOVA analysis of the time taken to solve the pre-test puzzle yielded a significant difference between the two groups \[F(1,185) = 5.972, p = 0.015\]: experimental group solved the puzzle significantly faster (312.4 ± 46.145 seconds at 95% confidence interval, N=62) than control group (436.73 ± 66.553 seconds, N=124). So, whereas the control and experimental groups were comparable in terms of prior preparation, experimental group solved the pre-test puzzle significantly faster than control group.

Mixed factor ANOVA analysis with pre-post as within-subjects factor and treatment (without versus with motivational support) as between-subjects factor yielded:

- Significant main effect for pre-post \[F(1,134) = 500.583, p < 0.001\]: the score for the two groups combined improved from 0.2708 ± 0.053 on the pre-test to 0.9495 ± 0.0185 on the post-test. So, the tutor was effective at helping students learn to solve Parsons puzzles.

- Significant main effect for treatment \[F(1,134) = 1569.136, p < 0.001\]: experimental group scored higher than control group on both the pre-test (0.3368 ± 0.0995 versus 0.2392 ± 0.0635 for control) and the post-test (0.9530 ± 0.2816 versus 0.9478 ± 0.1937 for control), although the difference between the two groups on the pre-test was not itself statistically significant.

But, the interaction between pre-post and treatment was not significant. So, the introduction of motivational supports did not affect pre-post improvement, and therefore, the amount of learning.

Similar mixed factor ANOVA analysis of the time spent per puzzle yielded:

- Significant main effect for pre-post \[F(1,133) = 150.639, p < 0.001\]: the time spent by the two groups combined decreased from 357.25 ± 36.205 seconds on the pre-test to 96.69 ± 11.322 seconds on the post-test. So, after practicing with the tutor, students were able to solve the puzzle significantly faster.

- Significant main effect for treatment \[F(1,133) = 440.564, p < 0.001\]: experimental group solved the pre-test puzzle a lot faster (281 ± 35.14 seconds) than control group (394.12 ± 49.32 seconds), thereby averaging faster times on pre- and post-test puzzles combined.

- The interaction between pre-post and treatment was significant \[F(1,133) = 11.061, p = 0.001\]: the pre-post improvement was greater for control group (from 394.12 ± 49.32 to 91.51 ± 13.62 seconds) than experimental group (from 281 ± 35.14 to 107.42 ± 20.18 seconds). While experimental group students would have spent a few seconds to read the screen that displayed the progress of the student after each puzzle, that alone could not have accounted for the 16 additional seconds experimental group students took on the post-test puzzle than control group students. Since students were made aware upfront of the minimal number of actions needed to solve each puzzle, and all their actions were enumerated, these motivational supports may have had the adverse effect of making experimental students more cautious, and hence, slower when solving the puzzles. This hypothesis bears further testing.

One of the feedback comments provided by students in fall 2015-spring 2016 was that they had to solve too many puzzles during the adaptive practice session. So, we reduced the mastery criteria used for adaptive practice: whereas students had to solve at
least 2 puzzles and score at least 80% in fall 2015-spring 2016, in fall 2016 they had to solve at least one puzzle and score at least 60%.

Univariate ANOVA analysis with the number of practice puzzles solved as the dependent variable and treatment as the fixed factor yielded a significant main effect for treatment \(F(1,169) = 3.774, p = 0.054\): control group solved 7.95 ± 1.298 puzzles during adaptive practice whereas experimental group with reduced mastery criteria solved 5.89 ± 1.218 puzzles. So, students solved fewer practice puzzles with reduced mastery learning criteria, but this was to be expected.

Univariate ANOVA analysis of the score per practice puzzle yielded no significant main effect for treatment: control group scored 0.7948 ± 0.0232 whereas experimental group scored 0.7657 ± 0.0413 per practice puzzle. However, based on the assumption that learning improves with the number of puzzles solved, we re-ran ANCOVA analysis with the number of practice puzzles solved as a covariate. The main effect for treatment was now significant \(F(2,169) = 3.19, p = 0.044\). So, after accounting for the fewer puzzles solved by the experimental group based on reduced mastery learning criteria, experimental group still scored less per puzzle during adaptive practice than control group.

Similarly, univariate ANOVA analysis of the time spent per practice puzzle yielded no significant main effect for treatment: control group spent 158.258 ± 15.626 seconds per puzzle whereas test group spent 165.71 ± 32.96 seconds. When we re-ran the analysis with the number of practice puzzles solved as a covariate, the main effect was significant \(F(2,167) = 6.522, p = 0.002\). So, after accounting for the fewer practice puzzles solved by the experimental group, experimental group spent more time per puzzle than control group. However, the difference of 7 seconds between the two group means can be explained as the time spent by experimental group viewing the progress screen displayed after each puzzle.

We plan to repeat this evaluation with larger sample sizes and better-matched student groups.

Acknowledgments. Partial support for this work was provided by the National Science Foundation under grants DUE-1502564 and DUE-1432190.

3 References