The Effect of Student Model on Learning

Adrian MARIES, Amruth KUMAR
Ramapo College of New Jersey
{amaries, amruth}@ramapo.edu

Abstract

Our goal in this study was to compare the effectiveness of displaying the open student model as a set of skillometers versus concept maps. The data suggests that concept maps are significantly more effective than a set of skillometers when answering questions that require synthesizing an overview of the topic.

1. Introduction

The constructivist theory of learning argues that we gain knowledge by building upon already existing knowledge, that we learn new concepts by integrating them with concepts we already know. Constructing concept maps makes students analyze the structure of their own knowledge which helps them assimilate the new information [9]. Several studies have shown that concept maps help students learn. One study found that creating concept maps helps students understand and retain the material presented in class [10,11] Another study found that studying the already worked-out concept map was more effective than generating the concept map from scratch, which was in turn more effective than generating the concept map from just a list of concepts or from a list that was already arranged spatially [12].

Several studies have shown the benefits of open student models. Making the student model available to students can make them aware of their own knowledge, or lack of it, and, in turn, improve their learning [3,4]. One survey shows that students want to have access to their learner models [1]. Student model presented in a table format is difficult to understand [6]. Therefore, visualization of data is a critical part of the open student model [7]. Concept maps, with their ability to reveal the relationships among concepts, have been proposed as a mechanism to present the learner model [1,2,5,8].

In our software tutors, we present the open student model as a concept map. We wanted to find out whether using the concept map to present the open student model conferred any benefits over using other techniques for visualizing the student model. In this paper, we will describe an experiment that we conducted to answer this question, analyze the collected data and present our results.

2. The Tutor and the Protocol

We used a software tutor on arithmetic expressions for our evaluation (www.problets.org). It presents problems on evaluating arithmetic expressions in C/C++/Java/C# (e.g., , 5 + 4 % 8), and asks the student to evaluate them step-by-step, i.e., one operator at a time. Once the student has entered his/her answer, the tutor provides delayed feedback – it lists how many steps the student solved correctly. It displays the correct evaluation of the expression using under-braces and intermediate results. In addition, it prints a text explanation for each step, such as “16 / 5 returns 3. Since both the operands are integers, integer division is performed. Any fraction in the result is discarded.” Since the student’s attempt is displayed in the left panel and the correct evaluation is displayed in the right panel simultaneously, the student can compare the two solutions.

In fall 2006 and spring 2007, we used the arithmetic expressions tutor to evaluate whether using the concept map as the student model conferred any additional benefits. The subjects were students in Psychology courses who used the tutor over the web, on their own time, as part of the experiential learning requirements of a Psychology course.

We used a controlled study and the traditional pre-test-practice-post-test protocol:

1. First, the subjects answered an online pre-test consisting of 9 multiple-choice questions. 6 of these were related to arithmetic expressions and 3 were unrelated.
2. The subjects worked with the software tutor for 15 minutes solving expression evaluation problems and reading the feedback. After solving each problem, the subjects were shown their open student model. For the test group, the model was
shown as a taxonomic concept map (domain concepts are nodes, links are is-a and part-of relations, and the map is an and-or tree), with their percentage completion of each concept graphically displayed in the corresponding node (See Figure 1). For the control group, the model was shown using skillometers (a series of progress bars graphically showing the completion percentage of each concept) (See Figure 2).

3. Finally, the subjects answered an online post-test consisting of the same questions as on the pre-test.

The questions on the pre-test and post-test were:

1. How many arithmetic operators are available in C++ programming language? 1/2/3/4/5/6. The correct answer was 5.
2. Pick ALL the operators among the following that are C++ arithmetic operators (check all that apply): <, *, !, /, %, ^. The answer was * and %.
3. How many of the following numbers are prime: 2, 3, 4, 5, 6, 7, 8, 9, 10. The answer was 4.
4. Which of the following C++ operators have 'integer' and 'real' types (check all that apply)? - <=, >, &&, +, /, ^. The answer was /.
5. For which of the following C++ operators is 'Dividing by Zero' an issue (check all that apply)? - >=, ||, %, ^, !. The answer was %.
6. What is the sum of the internal angles of a triangle? 90/180/270/360/450/600. The answer was 180.
7. To how many C++ arithmetic operators does the issue of 'Precedence' apply? – none/only one/all/only two/only three. The answer was all.
8. Pick all the issues that apply to all the C++ arithmetic operators (check all that apply) – Coercion, Correct evaluation, Error, Associativity, Real. The answer was Correct evaluation and Associativity.
9. Which of the following are types of operators in C++ (check all that apply)? – Relational, Abstraction, Repetition, Logical, Selection, Assignment. The answer was Relational, Logical and Assignment.

Note that almost all the questions had multiple correct answering options and the student was asked to select all those options. Questions 3, 6 and 9 are not related to arithmetic expressions, and were meant to serve as control questions for each subject. Answers to questions 1, 2, 4, 5, 7, and 8 cannot be synthesized without an overview of the domain of arithmetic operators, since these questions are not about any particular operator, but rather about groups of operators.

During the problem-solving session, the tutor never explicitly provided the answers to any of these questions. But, the answers to questions 1, 2, 4, 5, 7, and 8 were evident from examination of the open student models. The spatial organization of the concept map made these answers more obvious than the list organization of the skillometers. Take, for example, question 2: Pick ALL the operators among the following that are C++ arithmetic operators (check all that apply): <, *, !, /, %, ^. A quick look at the open student model displayed as a concept map, Figure 1, will show that the five C++ operators are the children of the root node, namely +, -, *, / and %. Even though this information is available in the student model displayed as skillometers as well, Figure 2, it is more difficult to find. Not only do you have to look at most topics in the skillometers, but the notation is not as straightforward either.

4. Data Analysis

50 students participated in the evaluation, 32 were in the test group and 18 in the control group. We used two grading schemes:

1. In the regular grading scheme, if a problem had $n$ answering options, $m$ of which were correct, the student received $1/m$ points for each correct answer. E.g., question 8 has 5 options, 2 of which are correct; if the students’ answer includes one
incorrect option and the two correct ones, they receive full credit.

2. In the negative grading scheme, students were penalized for guessing. If a problem had $n$ answering options, $m$ of which were correct, the student received $1/m$ points for each correct answer and lost $1/(n - m)$ points for each incorrect answer. E.g., if the students have the same answer as before to question 8, they only receive $2*1/2 - 1/3 = 0.66$ points.

**Aggregate of all the Questions.** First, we considered the aggregate of all the questions for each student. We conducted a 2 X 2 mixed factor ANOVA analysis of the aggregate scores with pre-post as the within-subjects factor and treatment (skillometers versus concept maps) as the between-subjects factor. Using regular grading, we found a significant main effect for time (pre versus post-test) $F(1,48) = 7.728, p = 0.034$ - students scored significantly higher on the post-test (4.879 points) than on the pre-test (4.158 points). There was no significant main effect for the treatment, or significant interaction between pre-post and treatment. We again found a significant main effect for time using negative grading $F(1,48) = 17.417, p = 0.000$ - students scored significantly more on the post-test (3.993 points) than on the pre-test (2.822 points). There was no significant interaction between pre-post and treatment.

**Related and Unrelated Questions.** We did a 2 X 2 X 2 mixed factor ANOVA analysis with pre-post scores and related-unrelated questions as within-subjects factors and treatment (skillometer versus concept map) as between-subjects factor. We found that:

- There was a significant main effect for related (ave 2.076) versus unrelated questions (ave 1.410) $F(1,48) = 16.023, p = 0.000$.
- There was a significant main effect for pre-test (ave 1.478) versus post-test (ave 2.008) $F(1,48) = 17.417, p = 0.000$ - a clear pre-post increase.
- There was significant interaction between related versus unrelated questions and pre-post test $F(1,48) = 24.769, p = 0.000$. The average on related questions increased from 1.525 on pre-test to 2.627 on post-test for related questions and decreased from 1.430 to 1.390 on unrelated questions.

We did not observe any other significant interaction.

**Related Questions.** Next, we repeated the 2 X 2 mixed factor ANOVA analysis for the aggregate scores on related questions only. Once again, there was a significant main effect for time using regular grading $F(1,48) = 7.833, p = 0.007$ - students scored significantly more points (3.235) on the post-test than on the pre-test (2.448). We found the same results with negative grading $F(1,48) = 23.289, p = 0.000$ - students scored significantly more points (2.654) on the post-test than on the pre-test (1.417). The interaction between pre-post and treatment was significant whether we used regular grading $F(1,48) = 3.925, p = 0.053$ or negative grading $F(1,48) = 3.476, p = 0.068$.

**Unrelated Questions.** We did not find any significant main effect for time using regular grading $F(1,48) = 0.401, p = 0.53$ or negative grading $F(1,48) = 0.252, p = 0.618$.

**Easy, Intermediate and Hard Related Questions.** Next, we repeated the above analysis for easy (1 and 2), intermediate (4 and 5) and hard (7 and 8) questions considered together. The criterion we used to divide the problems into the three categories was the likelihood of finding the answer by taking a quick look at the open student model. Question 1 (How many arithmetic operators are available in C++ programming language?), for example, can be answered fairly easily. A quick look at the concept map will allow us to see that there are 5 second-level nodes that have operator names. To answer the hard questions, we need to carefully inspect the student model. In order to find the answer to question 7 (To how many C++ arithmetic operators does the issue of 'Precedence' apply?), we have to look at all third-level nodes and scan for nodes with the specified name, a task that requires more than just a quick glance at the concept map. Intermediate questions are somewhere in between, harder than what we call easy questions, but easier to answer than the hard ones.
For easy questions, there was no main effect for time or treatment and no significant interaction between the two.

For intermediate questions, there was a significant main effect for time \( [F(1, 48) = 17.936, p = 0.000] \) – the average score improved from 0.66 on the pre-test to 1.22 on the post-test. The effect for treatment was marginally significant \( [F(1,48) = 3.0, p = 0.09] \) – the test group scored significantly lower than the control group on the pre-test (0.515 versus 0.941). The interaction between time and treatment was not significant.

For hard questions, there was no significant main effect for time or treatment, but the interaction between the two was significant \( [F(1,48) = 4.147, p = 0.047] \): whereas the control group score decreased from pre-test to post-test (from 0.794 to 0.588), the test group score increased from pre-test to post-test (0.646 to 0.894). We have summarized the pre-post change in scores on the three types of questions, and the statistical significance of the control-test group difference in table 1.

Table 1: Analysis of Easy, Intermediate and Hard Questions

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Pre-post Change Easy</th>
<th>Inter</th>
<th>Hard</th>
<th>Between-subjects p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-negative grading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skillometers</td>
<td>Avg. 0.041</td>
<td>0.353</td>
<td>-</td>
<td>0.206</td>
</tr>
<tr>
<td></td>
<td>St. Dev. 0.644</td>
<td>0.786</td>
<td>0.751</td>
<td></td>
</tr>
<tr>
<td>Concept Map</td>
<td>Avg. 0.183</td>
<td>0.667</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td></td>
<td>St. Dev. 0.774</td>
<td>0.816</td>
<td>0.743</td>
<td></td>
</tr>
<tr>
<td>Between-subjects p-value</td>
<td>0.495</td>
<td>0.196</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>Negative grading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skillometers</td>
<td>Avg. 0.471</td>
<td>0.459</td>
<td>-</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>St. Dev. 0.736</td>
<td>0.783</td>
<td>0.523</td>
<td></td>
</tr>
<tr>
<td>Concept Map</td>
<td>Avg. 0.576</td>
<td>0.776</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td></td>
<td>St. Dev. 0.868</td>
<td>0.717</td>
<td>0.672</td>
<td></td>
</tr>
<tr>
<td>Between-subjects p-value</td>
<td>0.657</td>
<td>0.174</td>
<td>0.017</td>
<td></td>
</tr>
</tbody>
</table>

Hard Questions. The analysis on questions 7 and 8, taken separately, revealed the following results. For questions 7, using regular grading, we did not find a significant main effect for either time or treatment. Similarly, there was no main effect for time and treatment for questions 8, but the interaction between the two was marginally significant \( [F(1, 48) = 2.988, p = 0.090] \). While the average for the control group on this question dropped from 0.215 to 0.079, the average for the test group increased from 0.147 to 0.171. However, using negative grading, we found a marginally significant difference between the control and test groups on question 8, as shown in Table 2.

Table 2: Analysis of Questions 7 and 8

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Pre-post Change Question 7</th>
<th>Question 8</th>
<th>Between-subjects p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-negative grading</td>
<td>-0.118</td>
<td>-0.088</td>
<td>0.136</td>
</tr>
<tr>
<td>Skillometers</td>
<td>0.6</td>
<td>0.476</td>
<td></td>
</tr>
<tr>
<td>Concept Map</td>
<td>0.152</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>Between-subjects p-value</td>
<td>0.185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative grading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skillometers</td>
<td>-0.118</td>
<td>-0.136</td>
<td>0.136</td>
</tr>
<tr>
<td>Concept Map</td>
<td>0.152</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>Between-subjects p-value</td>
<td>0.073</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question 8 is one of the five multiple answer questions the students had to answer. They are the only type of questions on which guessing has a reasonable chance of improving the score. The reason why there is no significant difference between the two groups using the grading scheme that doesn’t use negative grading is due to the fact that guessing helps them find the correct answer and, since it is a hard questions, many students resorted to guessing. Choosing two correct options and two incorrect ones (out of a total of two correct and three incorrect options), for example, will give students full credit using regular grading while only giving them 0.33 points using the negative grading scheme.

5. Discussion

We found a significant improvement from pre-test to post-test on questions that are related to arithmetic expressions, whether regular grading or negative grading was used. On the other hand, the scores on the questions unrelated to arithmetic operators did not improve with time. This means that the improvement on the related questions did not occur by chance. It shows that going through the tutor helped students better answer the questions related to arithmetic expressions.

We argue that the pre/post-test improvement is not due to the problems the students solved working with the tutor, but due to the open student model. Recall that on most questions, students were asked to select all the applicable options, e.g., all the arithmetic operators. Given this overview nature of the questions, one needs an overview of the topic to glean the correct answer. Such an overview is provided by the open student model.
model, whether it is presented as a set of skillometers or as a concept map. Concept map version of open student model has the advantage over skillometers in that it clarifies the is-a/part-of relationships among the concepts. Alternatively, students may have constructed an overview answer by choosing only the operators on which they solved problems, but this explanation contravene Occam’s razor.

The difference between control and test groups grows statistically more significant as we go from easy to intermediate to hard questions for both grading schemes (Table 1). Clearly, as the difficulty of the questions increases, the gap between the ease of answering the questions using the two different means of displaying the student model widens. While it is relatively straightforward to answer the easy questions using either of the two ways of displaying the open student model, it is harder to answer the hard questions using skillometers than the concept map. In other words, students are less likely to implicitly learn the relationships among concepts using a set of skillometers than using concept maps.

The difference between control and test groups on question 8 is more significant with negative grading than with regular grading (Table 2). Clearly, at least some of the students guessed at least some of the answers; some of these guesses were correct, and others were incorrect. By penalizing guessing, negative grading brought the differences between control and test groups into sharper focus.

Our goal in this study was to compare the effectiveness of displaying the open student model as a set of skillometers versus concept maps. The data suggests that concept maps are significantly more effective than a set of skillometers when answering questions that require synthesizes an overview of the topic.

6. References


