AC 2012-4551: CHARACTERIZING STUDENTS HANDWRITTEN SELF-EXPLANATIONS

Mr. James Herold, University of California, Riverside

James Herold earned his B.S. in computer science at California Polytechnic State University, Pomona, in 2004. He is currently a Ph.D. student in computer science at the University of California, Riverside.

Dr. Thomas Stahovich, University of California, Riverside

Thomas Stahovich received a B.S. in mechanical engineering from the University of California, Berkeley, in 1988. He received a M.S. and Ph.D. in mechanical engineering from the Massachusetts Institute of Technology in 1990 and 1995, respectively. He is currently Chair and professor in the Mechanical Engineering Department at the University of California, Riverside.

Characterizing students' handwritten self-explanations

Abstract

Prior work has shown that self-explanation leads to greater learning gains and that those students who can clearly explain their solution process are more likely to generate good solutions. To our knowledge, no studies have measured the impact that self-explanation has on a student's solution process. In this work, we explore an unprecedented database containing digital records of the pen strokes from over 120 students' coursework over an entire quarter. This unique data allows for never before possible analyses of students' solution processes. In this paper, we compare the differences in class performance and solution processes between two groups of these students, one group that was required to supply handwritten self-explanations, another that was not. This comparison reveals that students who generate self-explanations along with their homework solutions perform better, learn core concepts better, and solve problems more like an expert than those students not required to generate self-explanation.

Introduction

Self-explanation is the process by which a student provides, in words, a summary of their own understanding. In related work, students have been asked to generate self-explanations of the steps of a worked-out example or the rationale behind the students' own solutions to a problem. These self-explanations serve a metacognitive purpose, allowing students to evaluate and monitor their own understanding of concepts and enabling them to guide their own learning process. We show here, as numerous other studies have demonstrated, that self-explanation positively impacts student performance. Additionally, we demonstrate, for the first time, the positive impact self-explanation has on a student's solution process.

We have conducted a large-scale study in which over 120 students from an undergraduate mechanical engineering course in statics were given LiveScribe™ digital pens. These pens serve the same function as a traditional ink pen, but additionally, they digitize the pen strokes and store them as sequences of time-stamped coordinates. Students from this course were asked to complete all coursework using the pens, resulting in a digital database containing millions of pen strokes. This course was divided into four discussion sections. One section was selected as the experimental group while another served as the control. The students in the former were asked to respond to a number of self-explanation prompts for five of their nine homework assignments, while students in the latter were not asked to provide self-explanation. Additionally, we present performance results on Steif's concept inventory[1] for the two groups.

A comparison of homework performance for the two groups demonstrated the expected result that students who generate self-explanations performed significantly better on their homework assignments than those who did not. Similarly, a comparison of the performance on the statics concept inventory showed that the experimental group had significantly greater learning gains for the fundamental statics concepts than did the control group.

While improvements in learning gains are an important result, the unique character of our data set – time-stamped pen strokes – enables a much richer analysis of student performance. In particular, it enables us to examine the process by which a student completes the problems in an assignment. In the current work, we examine the order in which students solve the problems in an assignment.

To ground our analysis, we asked three experts to use digital pens to complete some of the same homework assignments our experimental and control group students completed. We then used statistical analysis techniques to compare the work from the control and experimental groups to that of the experts. This unique form of educational informatics is enabled by our novel database of student work. This analysis revealed that students who generated self-explanations solved problems more like the experts than did the students in the control group.

We begin with a survey of related work on metacognition as well as other self-explanation studies in the Related Work section. We then present an in-depth explanation of the user study and data analysis we performed in the Experimental Design section, followed by a presentation of our findings in the Results section, and lastly a discussion of those findings in the Conclusion section.

Related Work

Chi et al.[2] argue that "the metacognitive component of training is important in that it allows students to understand and take control of their learning process." Metacognition is the awareness of one's own learning process and it serves as a major foundation for research performed on self-explanation.

Mayer[3] used metacognition as a context for examining the differences between retention and transfer. The former is the application of knowledge from one problem to an identical problem, while the latter is the application of that knowledge to a different problem. Mayer argues that metaskill is an essential part of transfer. Whereas metacognition is the awareness of one's own cognitive processes, metaskill is the ability to control and monitor those processes. Metaskill strategies may be taught just as any other skill, such as arithmetic, via strategy instruction. For example, students who are taught basic reading skills as well as strategies for summarizing their own reading, perform better on transfer questions.[4] These results demonstrate the inadequacy of teaching only basic skills and the need to complement them with metacognitive skills. In this context, we use self-explanation as a means to develop metaskills.

Numerous studies have demonstrated the positive impact self-explanation has on student performance. Bielaczyc et al.[5] studied the impact of different self-explanation strategies on a student's ability to learn LISP programming. In their experiment, students were given instruction via an intelligent tutoring system. Some students were also trained to ask themselves a series of questions regarding their own understanding of worked out examples they viewed with the tutorial. The experiment revealed a significant difference between the learning gains from the pre- to posttest between students that did and did not generate self-explanation. In this study students provided self-explanation after viewing study materials but before solving problems. This differs from our study in which students generate self-explanation throughout their solution process.

Chi et al.[6] stated that, “generating explanations to oneself facilitates the integration of new knowledge.” To verify this statement, the authors conducted a study in which eighth grade students were asked to provide self-explanation as they read passages from a text on the circulatory system. This demonstrates that students who generated self-explanation performed significantly better than those who did not. This study differs from ours in that students explained passages they read, whereas in our study, students explained their own solution processes.

Chi et. al.[2] made comparisons between two groups of students: “poor” and “good” performing students. These students were asked to generate self-explanation after studying worked out example problems. The results of this study demonstrated that students who perform poorly are typically unable to generate sufficient self-explanation of the worked out example problems.

Weerasinghe and Mitrovic[7] investigated the impact that self-explanation, paired with the use of an intelligent tutor, has on student performance in a database design course. In the study, students in the experimental group were prompted for self-explanation by the tutoring system whenever the student made a mistake. This protocol was used as the authors claim that prompting students to explain most of their problem steps would be “too burdensome,” although no evidence for this is provided. Because there was no statistical analysis, the results were inconclusive.

Hall and Vance[8] investigate the impact that self-explanation has on student performance as well as self-efficacy in a statistics course. Students in the experimental group collaboratively solved problems in teams of three, providing self-explanations of the reasoning behind their answers to one another. Students in a control group solved the same problems individually. This study showed that students who generated collaborative self-explanation perform significantly better at solving problems than students who did not.

What these studies have in common is their use of summative performance assessments to show the positive impact that self-explanation has on learning gains. To our knowledge, prior work has not used automated, formative assessments which capture changes in students’ solution behavior.

Experimental Design

In the winter quarter of 2011, we conducted a study in which 132 students enrolled in an undergraduate mechanical engineering course on statics were given LiveScribe™ digital pens which they used to complete their homework, quizzes, and exams. This produced a digital record containing the time-stamped coordinates of the points on every pen stroke of their work. In total, we collected data from seven homework assignments, seven quizzes, two midterms, and the final exam.

All students enrolled in this course attended the same lectures, but were divided into four different discussion sections. During the discussions, supplemental lecture was provided and sample problems were solved on the board. One of the four discussion sections, which we call the self-explanation or SE group, was given a supplemental set of prompts with five of their nine homework assignments. In responding to these prompts, students generate a self-explanation of the reasoning behind each major step of their solutions. Students hand-wrote their explanations with their digital pens and submitted their responses along with their homework solutions. The following self-explanation prompts, which are from the third homework assignment (Figure 1 shows one of the problems from this assignment), are typical of the prompts we used:

1 Why did you select the system that you used for your free-body diagram?

2 Could you have selected some other system and still solved the problem?

3 How did you model each of the reaction forces? For example, did you consider the reaction to be a pivot, roller, contact with friction, etc.?

4 When computing moments for the moment equilibrium equation, why did you choose the particular point that you used to compute moments about? For example, if you computed moments about point A, why did you pick A and not some other point?

5 Could you have simplified the analysis by picking some other point to take moments about?

6 Why did you choose to solve the equilibrium equations in the order that you did? For example, if you solved the x-equilibrium equation, then the y-equilibrium equation, and finally the moment equilibrium equation, why did you chose this order?

1-The pin at A can support a maximum force of 3.2 kN. What is the corresponding maximum load L which can be supported by the brackets?

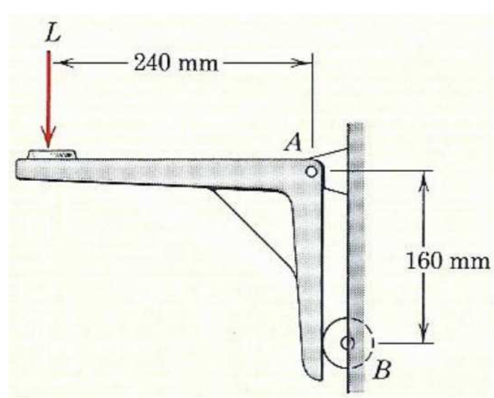


Figure 1: Homework 3 problem 1.

We elicited self-explanations only for those homework assignments that dealt primarily

HW No.	No. of Prob.	No. SE Students	No. SE Prompts	No. SO Students
3	8	26	6	31
4	7	21	6	29
5	8	25	4	31
6	6	23	4	29
8	5	22	3	30

Table 1: Details of the homework assignments including: the number of problems in the assignment, the number of self-explanation prompts, and the number of students in the SE and SO groups that completed the assignment.

with equilibrium. We excluded, for example, the first two assignments which covered prerequisite topics such as vectors and moments. Students from the SE group generated self-explanation for homework assignments three, four, five, six, and eight. Table 1 lists the number of experimental subjects and problems considered in the study.

The problems in each assignment followed a pattern in which the first three problems of the assignment were similar to each other and differed only superficially. The problems involved the same concepts, but the shapes of the objects varied. The second three problems of an assignment were also similar to each other. The remaining problems on each assignment were unique.

We used only one of the four discussion sections as a control. The remaining two sections were not used because one was an experimental group for another study, and the other contained only a small number of students who added the class late. For a variety of reasons, students of the latter group may have differed from the students in the other sections. We refer to the students of the control group as the "solution only" or SO group, as the students in this section did not generate self-explanations.

All students completed Steif's statics concept inventory[1] at both the beginning and end of the quarter. This is a multiple choice quiz which probes students' understanding of nine core statics concepts: free body diagrams, Newton's third law, statics equations, rollers, slots, negative friction, representation, friction, and equilibrium. There were three multiple choice questions for each of these concepts.

In addition to comparing two groups of students to each other, we established an expert protocol representing the solution sequence we would expect a student with an expert-stance on statics concepts to follow. To obtain this protocol, we asked three experts to solve some of the homework assignments and to respond to the same self-explanation prompts. These experts comprised one graduate and two undergraduate mechanical engineering students. The latter two had solved these same homework problems in their statics course two years prior. We found that these experts always solved their homework problems in sequential order. That is, experts completed problem one, then problem two, then three, and so on. They never solved problems out of order by skipping ahead or backtracking to a previous problem.

Results

Our first two analyses consider student performance on both homework grades and Steif's concept inventory.[1] We performed repeated measures analysis of variance (ANOVA) on both the homework and concept inventory performance. A number of the students had incomplete records. Some failed to complete all of the homework assignments and some failed to complete both the pre- and posttest concept inventory. As a remedy, we performed a missing values analysis to estimate missing homework and concept inventory scores. This technique estimates missing values by regressing on known values. The ANOVA results presented in both the Concept Inventory Analysis and Grade Analysis sections were done using these estimated missing values.

Our second two analyses consider the order in which students completed the problems in an assignment. We found that the experts always solved the problems in sequential order. Consequently, our analysis examines the extent to which the students solve problems out of order, an indication of novice behavior. To perform this analysis, we employ two sequence analysis methods commonly used in such disciplines as natural language processing and bioinformatics.[9–11]

Concept Inventory Analysis

The average number of questions students correctly answered on the concept inventory is shown in Figure 2. ANOVA revealed that the difference in the pre- to posttest learning gains between the two groups is significant ($p = 0.011$).

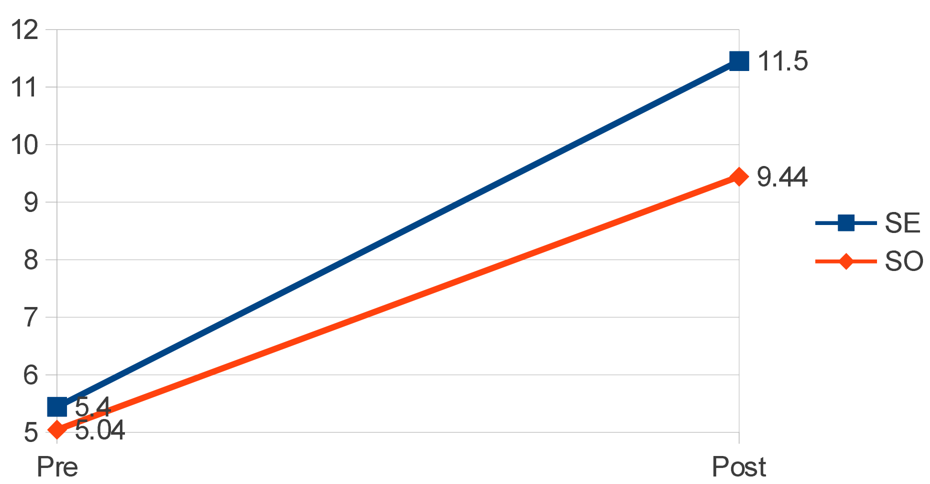


Figure 2: Pre/posttest scores for the SE (blue) and SO (orange) groups on the concept inventory.

Grade Analysis

Next, we compared the homework performance of the two groups. The average scores for each homework assignment are shown in Figure 3. It is important to note that in this course, the grade on a homework assignment was determined by the performance on one problem – the other problems were not graded. ANOVA revealed that the differences between the homework grades of the two groups is significant. More specifically, there is a significant difference ($p < 0.01$) in the slopes of the linear best fit lines for the homework scores as shown in Figure 3.

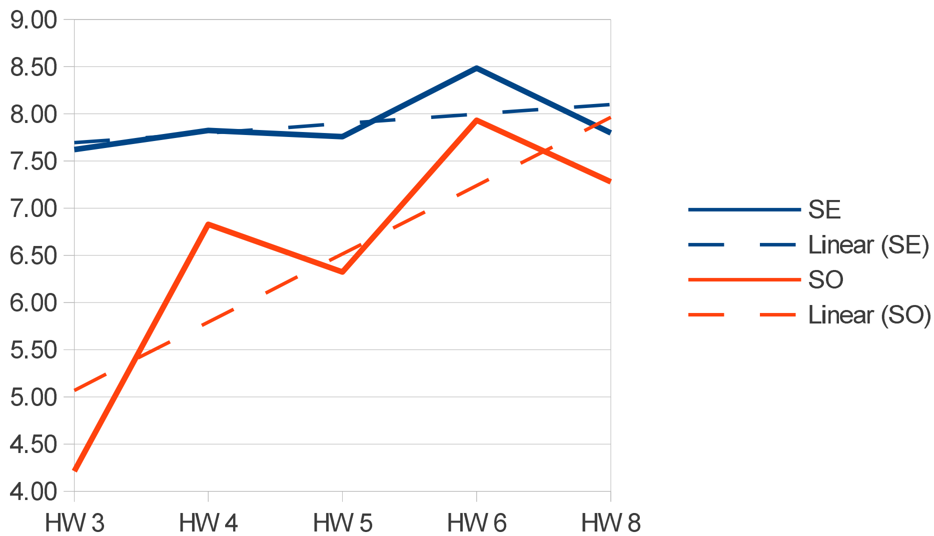
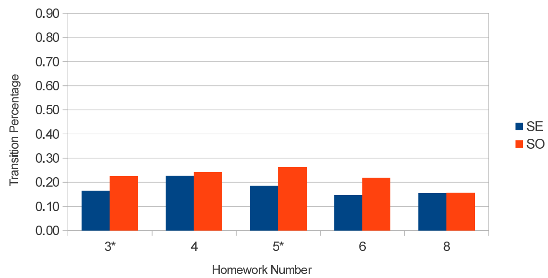


Figure 3: Average score for each homework assignment of both the SE (blue) and SO (orange) groups. The dashed lines represent the linear best fit of each. ANOVA revealed a significant difference in the slope of these two lines.

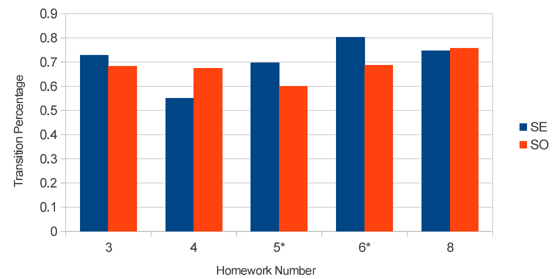
Transition Probability Analysis

To examine the order in which students solve the problems in an assignment, we represent their work as a sequence of problem numbers. For example, if a student begins with problem one, moves on to problem two, and then returns to problem one before working on problem three, the sequence would be "(1, 2, 1, 3)". In this example, there are two out-of-order problem transitions: 2-1 and 1-3. In our analysis, we consider three types of transitions: *in-order* – a transition to the immediately next problem, such as from 3 to 4; *skip* – a transition to a future problem, such as from 3 to 5; and *backtrack* – a transition to any earlier problem, such as from 3 to 2.

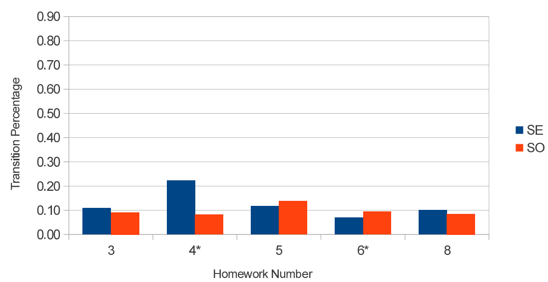
We compute the occurrences of each of these kinds of transitions for each student and normalize by the total number of transitions, yielding a transition frequency. Figure 4 shows the average transition frequencies on each homework assignment for both the SE and SO students. We used a t-test to determine if the differences between transition frequencies for the two groups are significant. The problems for which the differences are significant are indicated in the figure.



(a) Backtrack



(b) In-Order



(c) Skip

Figure 4: Average transition frequencies for the SE and SO groups for each homework assignment. An asterisk (*) next to the homework number indicates that the difference between the two groups is significant ($p < 0.1$) as determined by a t-test.

N-gram Analysis

Whereas the previous analysis considered the frequency of out-of-order transitions, here we consider the frequency of transitions between particular problems, using an N-gram analysis. An N-gram is a subsequence of length N taken from a student's problem number sequence. For example, a two-gram or bigram is a subsequence of length two and a three-gram or trigram is a subsequence of length three. Consider, for instance, the student problem number sequence (1, 2, 3, 2, 4) which contains 4 bigrams: (1, 2), (2, 3), (3, 2), and (2, 4), and 3 trigrams: (1, 2, 3), (2, 3, 2), and (3, 2, 4).

Our analysis focuses on the likelihood that the elements in the N-gram occur together. First we consider Dice's coefficient, which is defined only for bigrams. Consider two sets of bigrams: the set of bigrams in which a particular problem number, p_1, is the first element of each bigram and another set of bigrams in which some other problem number, p_2, is the second element of each bigram. Dice's coefficient provides a measure of "similarity" for these two sets, computed as:

$$S = \frac{2|X \cup Y|}{|X| + |Y|} \tag{1}$$

Here, $|X|$ is the number of times some problem, p_1, appears as the first element in a bigram, $|Y|$ is the number of times a second problem number, p_2, appears as the second element in a bigram, and $|X \cup Y|$ is the number of times the two problems appear in the same bigram, (p_1, p_2). S is a number between 0 and 1; the closer it is to unity, the greater the similarity between the two sets, or in other words, the more likely it is that the two problem numbers appear together.

We created two corpora, one containing every problem sequence from the SE group and one with all sequences from the SO group. We computed Dice's coefficient for each bigram in these corpora separately. The differences between the Dice's coefficients of the SE and SO groups are shown in Figure 5. Here, we compare the Dice coefficients only for bigrams whose problem numbers were out of order. A negative value indicates that the problem numbers in a bigram appear more frequently together in the SO corpus than in the SE corpus. Note that we do not present results for bigrams that occur fewer than five times as Dice's coefficients for such cases would be unreliable.

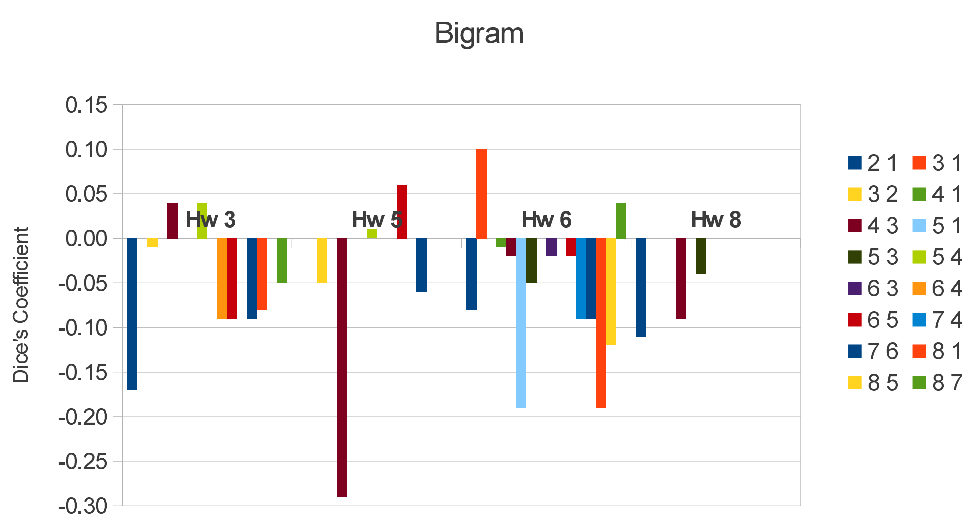


Figure 5: The difference in Dice coefficients (SE - SO) for bigrams occurring more than five times. Negative values indicate bigrams that are more similar in the SO than in the SE corpus.

Because Dice's coefficient is limited to bigrams, we consider mutual information when examining trigrams. This measures the statistical dependence of elements in a trigram. Mutual information can be thought of as the difference between the marginal entropy of a random variable, and the conditional entropy of that variable given a second:

$$I(X;Y) = H(X) - H(X|Y) \qquad (2)$$

$H(X)$ represents the uncertainty that some problem number, p_1, appears as the first element in any bigram and $H(X|Y)$ represents the uncertainty that, p_1 is the first element of a bigram given the occurrence of some particular second element of that bigram, say p_2. In this case, $I(X;Y)$ gives an indication of the amount of information gained from knowing that p_2 follows p_1. This calculation naturally generalizes to three variables, in which it is a measure of the amount of information gained about p_1 knowing that p_2 and some other problem, p_3, occur after it. $I(X;Y)$ is also a number between 0 and 1 such that the closer it is to unity, the more those problems are dependent on each other.

We compute mutual information separately for the SE and SO corpora. The differences in mutual information between the two corpora for each trigram are shown in Figure 6. Trigrams which appear fewer than five times do not appear in these results. A negative value in Figure 6 indicates that a sequence of problem numbers appears more frequently in the SO corpus than in the SE corpus.

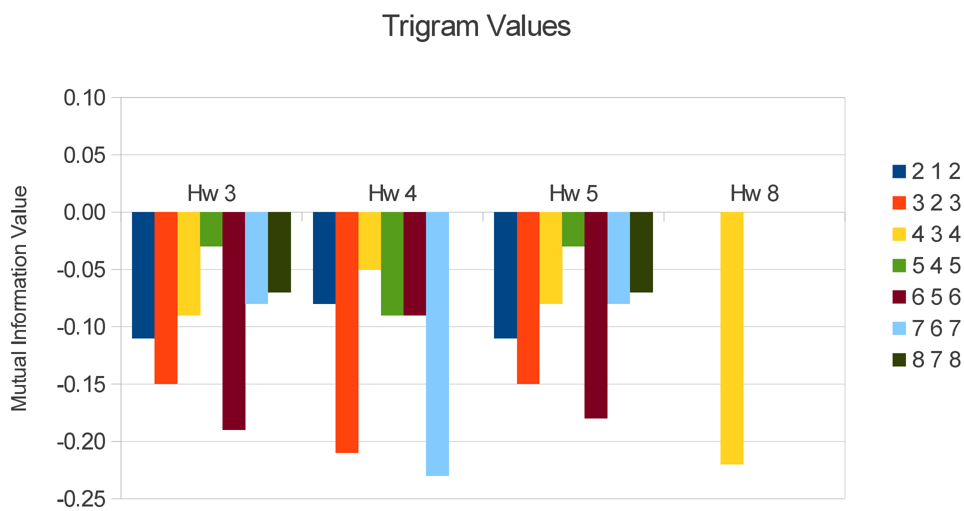


Figure 6: Difference in mutual information values between problem sequence trigrams of the SE and SO groups.

Discussion

Figure 2 shows that students in the SE and SO groups begin with equivalent conceptual understanding of statics concepts. The SE and SO groups on average got 5.4 and 5.04 questions right on the inventory pretest, respectively. At the end of the quarter, students from the SE group on average correctly answered 11.5 questions, while the SO group correctly answered on average 9.44. The difference in learning gains between the two groups is significant as revealed by ANOVA ($p = 0.011$). This evidence suggests that self-explanation may lead to greater conceptual understanding in this statics course. This corroborates intuition that self-explanation sharpens metaskills, leading to greater performance on transfer problems.

Figure 3 corroborates the well known story that self-explanation positively impacts performance. However, there is more to this story. Figure 3 shows that self-explanation can lead to a large boost in performance, but as time goes on, that difference dwindles to an insignificant level. The difference in the grades between the two groups is significant ($p < 0.1$) in the first three homework assignments, but not in the last two. This suggests that there may be a ceiling on student performance and that self-explanation, and the metaskills that it fosters, lead students to reach that ceiling quicker. This is an important benefit of self-explanation, especially in the context of a fast paced quarter system.

Our analysis of problem number sequences suggests that self-explanation can more quickly lead students to an expert-stance in the way that they solve problems. Both the transition, bigram, and trigram analyses show that students in the SE group typically solve problems out of order less frequently than the SO students. As indicated in Figure 4, whenever there is a significant difference between the transition frequencies of the two groups, SE students transition out of order less frequently. Similarly, Figure 5 shows that specific out-of-order bigrams appear more frequently for SO students than for SE student. Figure 6 presents perhaps the clearest distinction between the two groups; SO students always had more out-of-order trigrams than the SE students.

One possible explanation for this behavior may be related to the fact that the homework assignments contained groups of three consecutive problems that differed only superficially. It is possible that the SO students gained insights on subsequent problems, enabling them to revisit earlier problems in the set to correct their work. The SE students, on the other hand, may have better understood their work on the first attempt of a problem, making revisits to prior work unnecessary.

Conclusion

In this work we have compared, through a variety of analyses, the differences in performance and solution processes of students who did and did not generate self-explanation with their homework solutions.

In our first analysis, we compared the homework grades and pre- and posttest scores on the statics concept inventory. We found that students who generated self-explanation performed better on both the homework and the concept inventory posttest. This is an expected result, as prior work has demonstrated similar effects.

In our study, students used digital pens to complete all of their coursework. These pens enable us to create a digital record of every pen stroke on every problem. This unique digital database has enabled numerous novel analyses of student work. Here, we used sequential analysis techniques commonly used in bioinformatics and natural language processing to discover meaningful patterns in the order in which students solved homework problems. In doing so, we have demonstrated that self-explanation not only leads to greater performance gains, it also leads students to solve homework problems more like an expert. While these results are important, we believe that our unique database of student work contains a variety of other discoveries waiting to be made with other data mining techniques.

Acknowledgements

The authors gratefully acknowledge support for this work provided by the National Science Foundation via award No. 0935239.

References

[1] Paul S. Steif and John A. Dantzler. A statics concept inventory : development and psychometric analysis. *Journal of Engineering Education*, October 2005.

[2] T. H. Michelene; Miriam Bassok; Matthew W. Lewis; Peter Reimann; Robert Glaser Chi. Self-explanations: How students study and use examples in learning to solve problems. In *Cognitive Science*, pages 145–182, 1989.

[3] Richard E. Mayer. Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 1998.

[4] A.L. Brown and J.D. Day. Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 1983.

[5] Katerine Bielaczyc, Peter L. Pirolli, and Ann L. Brown. Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 1995.

[6] Michelene T. H. Chi, Nicholas De Leeuw, Mei hung Chiu, and Christian Lavancher. Eliciting self-explanations improves understanding. *Cognitive Science*, 18:439–477, 1994.

[7] Amali Weerasinghe and Antonija Mitrovic. Supporting self-explanation in an open-ended domain. In Mircea Negoita, Robert Howlett, and Lakhmi Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3213 of *Lecture Notes in Computer Science*, pages 306–313. Springer Berlin, 2004.

[8] Simin Hall and Eric A. Vance. Improving self-efficacy in statistics: Role of self-explanation and feedback. *Journal of Statistics Education*, 2010.

[9] Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. Class-based n-gram models of natural language. *Comput. Linguist.*, 18:467–479, December 1992.

[10] Andreas Stolcke. SRILM – an extensible language modeling toolkit. In *Proceedings of ICSLP*, volume 2, pages 901–904, Denver, USA, 2002.

[11] Hiroto Saigo, Jean-Philippe Vert, Nobuhisa Ueda, and Tatsuya Akutsu. Protein homology detection using string alignment kernels. *Bioinformatics*, 20:1682–1689, July 2004.