Learning by Problem Solving versus by Exmaples: The Benefits of Generating and Receiving Information

Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Lovett, Marsha C.

1992

Learning by Problem Solving versus by Examples: The Benefits of Generating and Receiving Information

Marsha C. Lovett*
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
lovett@psy.cmu.edu

Abstract

This experiment contrasts learning by solving problems with learning by studying examples, while attempting to control for the elaborations that accompany each solution step. Subjects were given different instructional materials for a set of probability problems. They were either provided with or asked to generate solutions, and they were either provided with or asked to create their own explanations for the solutions. Subjects were then tested on a set of related problems. Subjects in all four conditions exhibited good performance on the near transfer test problems. On the far transfer problems, however, subjects in two cells exhibited stronger performance: those solving and elaborating on their own and those receiving both solutions and elaborations from the experimenter. There also was an indication of a generation effect in the far transfer case, benefiting subjects who generated their own solutions. In addition, subjects' self-explanations on a particular concept were predictive of good performance on the corresponding subtask of the test problems.

Introduction

Solving problems and studying examples are both viable methods for learning to solve problems. Solving practice problems provides subjects with the experience of "doing" and forces them to consider all aspects of the solution. Studying example problems accompanied by their solutions (together, called worked examples) provides students not only with the correct answers but also with some information on how those answers were obtained. So, which method (if either) is better? Experimental results on that question do not yet provide a definitive answer.

Sweller and Cooper (1985) have found that subjects using worked examples required less study time and exhibited better near transfer performance than subjects solving problems. However, their method of learning by examples actually included problem solving half of the time.[1] In addition, many researchers have found that the particular content of worked examples can seriously affect subjects' learning outcomes (e.g. Catrambone, 1991; Pirolli, 1991; Ward & Sweller, 1990). Therefore, the experimental procedure used to implement learning by examples and the content of the examples are important variables to consider when evaluating this method. In the case of learning by problem solving, other variables (e.g. the type and timing of feedback) must be taken into account since they too have been shown to affect subsequent performance (e.g. Lewis & Anderson, 1985; Schooler & Anderson, 1990). Finally, individual differences among subjects can also influence the efficacy of these instructional methods. For example, the relationship between the quality of subjects' self-explanations[2] and subsequent performance has been demonstrated (Chi et al., 1989; Pirolli & Bielaczyc, 1989).

Experiment

The current experiment attempts a preliminary comparison between learning by examples and learning by problem solving, while dealing with many of these issues in a systematic way. Subjects worked on three practice problems in introductory probability. They were either provided with solutions or asked to generate their own, and they were either provided with complete, elaborate explanations of the solutions or asked to

---

*The author was supported by a NSF Graduate Fellowship. Funding for subjects was provided through MDR-87-51890 granted to Dr. John R. Anderson.

[1] For example, Experiment II contained four pairs of isomorphic practice problems. Subjects in the worked-example group studied a worked example for the first problem in each pair and then solved the second.

[2] Self-explanations are the elaborations generated by subjects, usually while studying an example problem, in which they explain various concepts to themselves.

create their own. This design allowed us to compare the two instructional methods *and* to manipulate the content of worked examples in the same experiment.

We denote these four conditions using the form <x>-<y>, where <x> corresponds to the source of the solutions and <y> corresponds to the source of the elaborations. The conditions can be characterized as follows: *subject-subject* = learning by problem solving and self-explaining; *experimenter-subject* = learning by studying "sparse" examples and self-explaining; *subject-experimenter* = learning by problem solving with explanatory feedback; and *experimenter-experimenter* = learning by studying elaborate worked examples. (See Appendix A for an example of the information provided to subjects in the four conditions.)

In accordance with previous results on the generation effect (e.g. Bobrow & Bower, 1969; Slamecka & Graf, 1978), we expected subjects who generated their own problem information to perform better. For example, generating one's own solution may make it more memorable and so improve subsequent performance. In the case of explanations, however, a generation effect may have to compete with an effect from high quality elaborations. Subjects creating their own elaborations may not perform as well as subjects who receive the experimenter's elaborations, if the latter are of substantially higher quality. Thus, we predicted a generation effect for solutions but not for elaborations.

Method

Subjects. Subjects included 50 undergraduate students at the University of California at Berkeley, who had little or no background in probability theory. All subjects were paid for their participation. Two subjects' data were removed from analysis: one because of inability to learn the material and one due to equipment failure resulting in incomplete data. This left 48 subjects in the experiment, twelve per group. Assignment to these groups was random.

Materials. Before solving the three practice problems, subjects were given a three-page introduction to the necessary concepts in probability. It covered the probability of an event E (defined as the number of outcomes satisfying E divided by the total number of possible outcomes) and the probability of multiple independent events (calculated by multiplying the probabilities of the individual events). The information in the introductory text prepared students to follow only one path to the solution of each problem, even though more than one was possible.

Eleven elementary probability problems were used in this experiment (including four adapted from Ross, 1989). All three practice problems were permutation problems with people choosing objects (e.g. scientists choosing from a pool of computers, see Appendix A). For the eight test problems, half were similar permutation problems (near transfer) and the other half were combination problems (far transfer). Combination problems differ from permutation problems in that the exact order of events does not matter. This difference affects the solution, mainly by changing the calculation of the numerator. In addition, both near and far transfer problems were split according to whether they contained people choosing objects or objects being assigned to people. Ross (1989) first found that these role assignments could affect performance: when an example problem, in which humans choose objects, was followed by role-reversed test problems, performance was worse than when the same roles were maintained between practice and test. (See Appendix B for sample test problems.) The test problems were given in the same order to all subjects -- from "nearest" transfer to "farthest".[3]

Procedure. In this experiment, subjects went through a lesson in probability. They read some introductory text, worked on three practice problems, and then solved eight test problems. All subjects were instructed to provide talk-aloud protocol during the practice and test problems and, when applicable, to read experimenter-provided elaborations out loud. Mistakes made by subjects solving the practice problems were treated as follows: in the subject-subject condition, only the fraction corresponding to the correct step (e.g. 1/11) was provided; and in the subject-experimenter condition, both the fraction and the prepared elaboration for that step were given.[4] Subjects never received coaching or corrections on their own elaborations. In order to equalize subjects' time-on-task for the practice problems, we ensured that all subjects spent approximately three minutes studying/solving each practice problem. Then, for the test problems, subjects were asked to work as quickly and accurately as possible.

---

[3] Subjects received the test problems in the following order: permutation problems with the same roles as the practice problems (people choosing objects), permutation problems with reversed roles (objects being assigned to people), combination problems with same roles, and finally, combination problems with reversed roles.

[4] Note that the manipulations in this experiment could not be purely executed such that subjects generated everything or nothing according to their condition. For example, 5 of the 12 subjects in the subject-subject cell received at least one correction from the experimenter. Nevertheless, cases like these only made the groups more similar and, hence, made differences between groups harder to find.

Results and Discussion

In analyzing these data, we were mainly interested in finding performance differences that might exist between the four experimental groups. We also wanted to explore two other questions about the data: Did subjects' performance on the test problems vary significantly according to transfer distance (near vs. far) and role correspondence (same vs. reversed) between the practice and test problems? And, in what way did individual differences impact on subjects' performance?

A 2x2x2x2 Mixed ANOVA on subjects' percentage of test problems solved correctly provided an overall analysis of the performance data. The between-subjects factors in this analysis were source of solutions and source of elaborations, and the within-subjects factors were transfer distance and role reversal. (See Table 1 for this breakdown of the data.) With respect to the between-subjects factors, no main effect for source of solution or for source of elaboration was found, but the interaction of these factors was significant ($F_{sol}(1,44) = 1.24$; $F_{elab}(1,44) < 1$; $F_{inter}(1,44) = 5.82, p < .05$; $MSE = 1510$). A post-hoc analysis of these data indicated that the interaction was due to high performance in the subject-subject and experimenter-experimenter groups compared to the other two ($F(1,46) = 6.33, p = .02$; $MSE = 375$).

Since this performance pattern arises in later analyses, it is worthwhile to consider it here. First, the finding that subjects who solve and explain on their own perform well is consistent with the generation effect as described above. However, the high performance of subjects who received all their problem-solving information from the experimenter shows the opposite of a generation effect. It seems that receiving correct solutions *and* high quality explanations benefited these subjects. Indeed, the experimenter-provided elaborations were generally of higher quality than subjects' own elaborations. For example, only five out of the 24 subjects elaborating for themselves verbalized three of the most important concepts found in the experimenter-provided elaborations. (Later, we will present evidence that the subjects with more complete self-explanations *do* perform better.)

	s-s	s-e	e-s	e-e	**all**
near-same	96	92	63	88	84
near-reversed	50	46	50	42	47
near overall	**73**	**69**	**54**	**65**	**66**
far-same	17	0	4	17	9
far-reversed	38	17	4	38	24
far overall	**27**	**8**	**4**	**27**	**17**
all test	50	39	29	46	41

Table 1: Average percentage of test problems correct, by experimental condition and test problem type

This interpretation leads to the prediction that subjects in the experimenter-subject cell should perform poorly because they lack the opportunity to generate solutions and they lack high quality elaborations. In fact, this cell did exhibit the worst performance. The subject-experimenter cell, however, did not perform quite as well as a generation effect and high quality elaborations would predict. One possible reason is the procedural awkwardness involved in this condition; subjects solved a step in the solution and then were asked to read the experimenter's elaboration for that step. Although subjects were forewarned about this procedure, it still might have interfered with their concentration and memory load. In fact, six subjects in the subject-experimenter cell required hints compared to only one in the subject-subject cell ($chi2 = 5.09, p < .05$), even though there is no reason to suspect ability differences existed between these two groups. Receiving hints more often could help explain the performance of the subject-experimenter subjects because each hint made them miss an opportunity to generate a solution step. A second explanation proposes that inconsistent information sources resulted in poor performance in the subject-experimenter and experimenter-subject cells because these subjects had to integrate the experimenter's information with their own. This increased cognitive load may weaken the subjects' problem memories. Unfortunately, the present data cannot tease apart these alternative explanations.

Also, in the 2x2x2x2 ANOVA mentioned above, we found some interesting within-subject effects. (See Table 1) Not surprisingly, subjects performed better on near transfer than far transfer test problems ($F(1,44) = 98.2, p < .001$). In addition, subjects performed better on test problems with the same roles as the practice problems, compared to test problems with reversed roles ($F(1,44) = 6.81, p = .01$). This replicates Ross's (1989) finding and suggests that subjects may be using analogical problem solving in this experiment as well. (See Comparison with Related Work.) The only other significant effect in this analysis was the transfer distance x role reversal interaction ($F(1,44) = 54.8, p < .001$). This interaction might have occurred because of learning during the test phase; in particular, subjects did surprisingly well on the very last test problem.

Near transfer

Since the near transfer and far transfer test problems resulted in different performance, we analyze them separately. The average percentage correct for near transfer problems is presented in Row 3 of Table 1. A 2x2 ANOVA (source of solutions x source of elaborations) of these data did not reveal any significant differences between the cells ($F_{sol}(1,44) = 2.32$, n.s.; all other F's < 1). Subjects in all four conditions seemed to do quite well.

Far transfer

The average percentage correct for far transfer problems is presented in Row 6 of Table 1. A 2x2 ANOVA (source of solutions x source of elaborations) on these data revealed no main effects, but it did reveal the same interaction as in the total performance measure ($F(1,44) = 7.94$, $p < .01$, $MSE = 656$). Namely, the subject-subject and experimenter-experimenter conditions performed the best ($F(1,46) = 8.27$, $p < .01$; $MSE = 630$). This pattern of results can be explained in the same way as the pattern for the total performance data. (In fact, the total performance differences are due in most part to these far transfer data, since the near transfer data did not differentiate much between the cells.)

More specific data are available with respect to far transfer performance. Recall that, in the far transfer problems, the numerator's starting value must be calculated differently than in the near transfer problems but that the other solution steps are similar to the near transfer problems. Therefore, looking at subjects' choice of numerator starting value (NSV) provides a sharper measure of far transfer performance. The total number of far transfer problems on which subjects had the correct NSV was: subject-subject 29; subject-experimenter 14; experimenter-subject 8; and experimenter-experimenter 19. A 2x2 ANOVA (source of solutions x source of elaborations) on these data revealed the same interaction found in other analyses ($F(1,44) = 8.14$, $p < .01$; $MSE = 1.73$) as well as a marginal main effect indicating that subjects who solved the practice problems performed better than those provided with solutions ($F(1,44) = 3.08$, $p = .08$; $MSE = 1.73$). Again, these data are indicative of a generation effect and a quality of explanations effect.

Individual data

One of the main features of this experiment is that it contains another experiment within its cells. By analyzing the protocols of the subjects in the experimenter-subject condition, we can look at the effects of individual subjects' self-explanations on subsequent performance. For example, we compared subjects' self-explanations of a particular concept in the practice problems with subsequent performance on corresponding parts of the test problems. We chose NSV for this analysis because it plays an important role in these problems (especially for far transfer), and it happened to be the only important concept that was differentially elaborated upon by subjects.[5] Six subjects self-explained NSV and six subjects did not. Of those who did, four ended up getting NSV correct on the test. Of the subjects who did *not* self-explain NSV, none ended up getting NSV correct on the test. Thus, specific self-explanations had a significant effect on specific performance ($chi2 = 6$, $p < .01$).

These results also provide support for the notion that the quality of explanations can affect performance. Above, subjects who self-explained NSV tended to get NSV correct on the test problems. When subjects received the experimenter's elaborations (which always included NSV), they also exhibited this tendency. The proportions of subjects who got NSV correct in at least one test problem are as follows: subject-experimenter 7/12; experimenter-experimenter 8/12. These two proportions are not significantly different from the proportion of subjects in the experimenter-subject cell (4/6) who explained NSV during practice and got it correct at least once during the test ($chi2 \sim 0$, n.s.). However, these proportions are different from the proportion of subjects in the experimenter-subject cell (0/6) who *did not* explain NSV during practice and then got it correct during the test ($chi2 = 7.5$, $p < .01$). These results suggest that (in three experimental conditions) subjects with an elaboration of NSV during practice tend to get the NSV correct at test, regardless of whether that elaboration was provided by the experimenter or self-explained.

For the subject-subject condition, we also evaluated the effectiveness of self-explaining NSV. Here, the proportion of subjects getting NSV correct on the test problems was higher than in the other conditions, regardless of whether the subjects self-explained about NSV during practice. Specifically, all three of the subjects who self-explained NSV during practice got it right during the test, and eight of the nine subjects who did not self-explain NSV during practice got it right during the test. The latter proportion is much greater than the proportion of subjects in the experimenter-subject cell (0/6) who did not self-explain NSV but still got it right ($chi2 = 11.43$, $p < .001$). This large difference indicates that subjects in the subject-subject condition were learning about NSV, whether or not they self-explained about it. This is an advantage of learning by problem solving that is not otherwise captured in our performance data.

Comparisons with Related Work

Research by Ross (1989) and Catrambone (1991) is especially relevant to this work because it uses the same type of probability problems in the context of learning

---

[5] Recall that, in the near transfer problems, the numerator is always 1, but in the far transfer problems, subjects must calculate the numerator starting value.

by studying worked examples. Ross (1989) found that human/object roles affected subjects' use of their example problem memories in such a way that reversing the human/object roles between the example and test problems made the test problems more difficult to solve. Likewise, Catrambone (1991) varied the human/object roles between example and test problems and produced a similar result. In the current experiment, we also found subjects' performance to be significantly worse on the role-reversed test problems than on the same-role test problems. Thus, our finding lends further support for Ross's (1989) conclusion that subjects may be using an object-mapping approach that is affected by human/object role correspondences.

Catrambone's (1991) experiment also resembles the current experiment because the content of worked examples was varied. Of Catrambone's four instructional groups, only two are comparable to the current design: the *numerator/denominator-subgoal* group in which subjects received an elaborated description of each step in the solution, including explanations for the choice of numerator and denominator, and the *subgoal* group in which subjects received only a brief description of each step in the solution. Catrambone's numerator/denominator-subgoal condition is virtually equivalent to our experimenter-experimenter condition. His subgoal condition, however, resides somewhere between our experiment-subject and experimenter-experimenter conditions in the amount of information it provides to subjects. Comparing percentages of test problems solved correctly by subjects in these conditions demonstrates substantial consistency between the two experiments. For example, Catrambone's numerator/denominator-subgoal and subgoal groups averaged 70 and 66 percent correct, respectively, on permutation problems (isomorphic to the example) and approximately 14 and 0 percent correct, respectively, on combination problems. These values are quite similar to those of our quasi-corresponding conditions (experimenter-experimenter and experimenter-subject): 65 and 54 percent correct for the permutation problems and 27 and 4 percent correct for the combination problems. In fact, the ordering of performance (across the two experiments) is fairly consistent with an ordering of information provided to subjects. In the current experiment, however, we embedded this comparison of worked examples with different contents in the larger context of comparing two instructional methods. We have also included the analysis of subjects' protocols within this same experiment. These features allowed us to capture differences between conditions and between individual subjects that might otherwise have been missed.

Conclusions

In this experiment, we found reliable performance differences between subjects generating and receiving different amounts of practice problem information. We found that on far transfer problems, subjects in two groups performed best -- those generating solutions and explanations for the practice problems on their own and those receiving high quality solutions and explanations from the experimenter. In addition, on a particular subtask of the test problems, subjects who solved the practice problems performed better than those who received solutions, regardless of their source of elaborations. These results support the existence of a generation effect which benefits subjects who solve problems on their own in the domain of probability.

We also found evidence that the quality of elaborations during practice can greatly improve subsequent performance. High performance in the experimenter-experimenter cell is one example of this since experimenter-provided elaborations were generally of higher quality than subjects' self-explanations. In addition, we examined the effects of individual subject's self-explanations and found that subjects who verbally elaborated on a particular concept in the practice problems performed better on corresponding steps in the test problems than subjects who did not.

Acknowledgements

I would like to thank Michael Ranney, Peter Pirolli, and the members of the Reasoning and CSM groups at the University of California at Berkeley, School of Education. I would also like to thank my advisor, John Anderson, for all his help and support.

Appendices

Appendix A: Sample practice problem and the information given in different conditions

The supply department at IBM has to make sure that scientists get computers. Today, they have 11 IBM computers and 8 scientists requesting computers. The scientists randomly choose their computer, but do so in alphabetical order. What is the probability that the first 3 scientists alphabetically will get the lowest, second lowest, and third lowest serial numbers, respectively, on their computers?

1/11 is the probability that the first scientist alphabetically will get the computer with the lowest serial number because there is only 1 computer with the lowest serial number and there are 11 computers for the first scientist to choose from.

[...]
To get the overall probability, we multiply the probability of each scientist choosing a particular computer:

ANSWER: 1/11 x 1/10 x 1/9

Note: The subject-subject condition received only the problem statement. The subject-experimenter condition received the problem statement and was asked to read each explanation (in italics) after solving the corresponding step. The experimenter-subject condition received only the problem statement and the answer. The experimenter-experimenter condition received the problem statement and all the elaborations (in italics).

Appendix B: Sample test problems

Near Transfer, Same Roles:
South Side High School has a vocational car mechanics class in which students repair cars. One day there are 12 students and 15 cars requiring repairs. The students randomly choose the cars, but go in order of their grades on the last mechanical exam (highest grade choosing first). What is the probability that the 6 cars in the worst shape are worked on by the 6 students with the highest grades on the last mechanical exam, in order of their grades? (i.e. the highest grade student working on the worst car, etc.)
[Answer: 1/15 x 1/14 x 1/13 x 1/12 x 1/11 x 1/10]

Near Transfer, Reversed Roles:
A group of 8 co-workers went to Pizza Hut to try the Personal Pan Pizzas. When they made their order (8 Personal Pan Pizzas with all the toppings), there were 10 such pizzas being taken out of the oven, one at a time. If Pizza Hut's policy is to serve their pizzas as soon as they come out of the oven and the food server distributes pizzas randomly to the co-workers, what is the probability that the first 5 pizzas to come out of the oven will be given to the 5 most senior co-workers, in order? (i.e. the first pizza going to the most senior, etc.)
[Answer: 1/8 x 1/7 x 1/6 x 1/5 x 1/4]

Far Transfer, Reversed Roles:
In the women's locker room at South Side High School, certain lockers are set aside for the 18 female swimmers on the school team. At the beginning of the season, these lockers are assigned to swimmers, starting with a row of 14 lockers next to the showers. These lockers are assigned at random, starting at the end near the showers and going down the row. What is the probability that the 7 lockers closest to the showers are assigned to the 7 butterfly swimmers?
[Ans.: 7/18 x 6/17 x 5/16 x 4/15 x 3/14 x 2/13 x 1/12]

References

Bobrow, S.A. & Bower, G.H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology, 80,* 455-461.

Catrambone, R. (1991). Helping learners acquire subgoals to improve transfer. *Cognitive Science Proceedings,* Chicago, Illinois. Hillsdale, New Jersey: Erlbaum.

Chi, M.T.H., Bassok, M., Lewis, M.W., Riemann, P., & Glaser, R. (1989). Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Lewis, M.W. & Anderson, J.R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. *Cognitive Psychology, 17,* 26-65.

Pirolli, P. (1991). Effects of examples and their explanations in a lesson on recursion: A production system analysis. *Cognition and Instruction, 8,* 207-259.

Pirolli, P. & Bielaczyc, K. (1989). Empirical analyses of self-explanation and transfer in learning to program. *Cognitive Science Proceedings,* Ann Arbor, Mich. Hillsdale, New Jersey: Erlbaum.

Ross, B.H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456-468.

Schooler, L.J. & Anderson, J.R. (1990). The disruptive potential of immediate feedback. *Cognitive Science Proceedings,* Cambridge, Mass. Hillsdale, New Jersey: Erlbaum.

Slamecka, N.J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4, 592-604.

Sweller, J. & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59-89.

Ward, M. & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7,* 1-39.